

PB85-240455

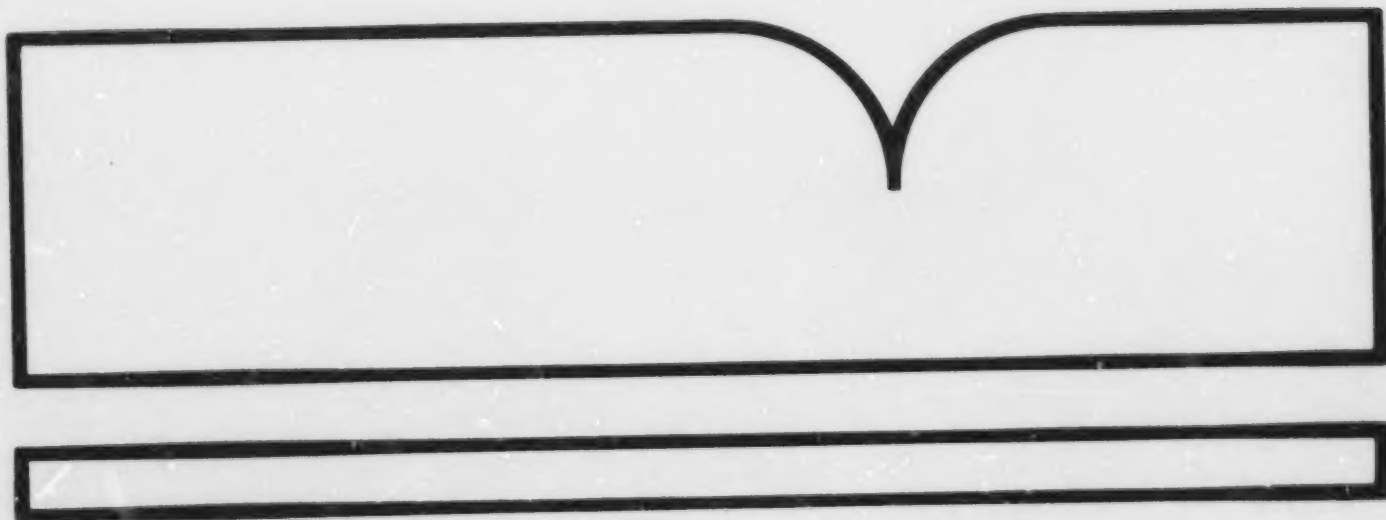
Methodology for Characterization of
Uncertainty in Exposure Assessments

Research Triangle Inst.
Research Triangle Park, NC

Prepared for

Environmental Protection Agency, Washington, DC

Aug 85



U.S. Department of Commerce
National Technical Information Service

NTIS.

EPA/600/8-85/009
August 1985

**METHODOLOGY FOR CHARACTERIZATION OF UNCERTAINTY
IN EXPOSURE ASSESSMENTS**

by

Roy W. Whitmore

**Research Triangle Institute
Research Triangle Park, NC 27709**

Contract No.: 68-01-6826

**Task Manager: James W. Falco
Project Officer: John Warren**

**OFFICE OF HEALTH AND ENVIRONMENTAL ASSESSMENT
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, DC 20460**

TECHNICAL REPORT DATA		
(Please read instructions on the reverse before completing)		
1. REPORT NO. EPA/600/8-85/009	2.	3. RECIPIENT'S ACCESSION NO. PB85 240455 iAS
4. TITLE AND SUBTITLE Methodology for Characterization of Uncertainty in Exposure Assessments	5. REPORT DATE August 1985	
	6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Roy W. Whitmore	8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Triangle Institute Research Triangle Park, NC 27709	10. PROGRAM ELEMENT NO.	
	11. CONTRACT/GRANT NO. Contract No. 68-01-6826	
12. SPONSORING AGENCY NAME AND ADDRESS Exposure Assessment Group (RD-689) Office of Health and Environmental Assessment U. S. Environmental Protection Agency Washington, D. C. 20460	13. TYPE OF REPORT AND PERIOD COVERED	
	14. SPONSORING AGENCY CODE EPA/600/21	
15. SUPPLEMENTARY NOTES Project Officer: James W. Falco, Exposure Assessment Group, U. S. Environmental Protection Agency, Washington, D. C. 20460		
16. ABSTRACT Exposure assessments, except those based upon measured exposure levels for a probability sample of population members, rely upon a model to predict exposure. The model may be any mathematical function that estimates the population distribution of exposure or an individual's exposure as a function of one or more input variables. Whenever a model that has not been validated is used as the basis for an exposure assessment, the uncertainty associated with the exposure assessment may be substantial. The primary characterization of uncertainty is at least partly qualitative in this case, i.e., it includes a description of the assumptions inherent in the model and their justification. Plausible alternative models should be discussed. Sensitivity of the exposure assessment to model formulation can be investigated by replicating the assessment for plausible alternative models. When an exposure assessment is based upon directly measured exposure levels for a probability sample of population members, uncertainty can be greatly reduced and described quantitatively. The primary sources of uncertainty are measurement errors and sampling errors. A quality assurance program should be designed into the study to ensure that the magnitude of measurement errors can be estimated. The effects of all sources of random error should be measured quantitatively.		
17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
18. DISTRIBUTION STATEMENT Release to the Public	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 95
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

DISCLAIMER

The information in this report has been funded wholly or in part by the United States Environmental Protection Agency under Contract No. 68-01-6826 to the Research Triangle Institute. It has been subject of the Agency's peer and administrative review, and it has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

	<u>Page</u>
LIST OF TABLES	iv
LIST OF FIGURES	~~iii~~ ✓
FOREWORD	(iv)
ABSTRACT	vii
ACKNOWLEDGMENTS	viii
1. INTRODUCTION AND OVERVIEW	1
1.1 What is an Exposure Assessment?	1
1.2 Characterization of Uncertainty	5
2. ASSESSMENTS BASED UPON LIMITED DATA	12
2.1 Limited Data for Directly Measured Exposures	12
2.2 Limited Data for Model Input Variables	13
2.2.1 Population Exposure Models	13
2.2.2 Personal Exposure Models	14
3. ASSESSMENTS BASED UPON ESTIMATION OF INPUT VARIABLE DISTRIBUTIONS	17
3.1 Estimation Based on Input Variable Distributions	17
3.2 Limiting Distributional Results	21
4. ASSESSMENTS BASED UPON DATA FOR MODEL INPUT VARIABLES	25
4.1 Interval Estimates of Exposure Percentiles: Input Variable Distributions Not Known	25
4.2 Interval Estimates of Exposure Percentiles: Input Variable Distributions Known	26
5. ASSESSMENTS BASED UPON EXPOSURE MEASUREMENTS.	29
6. AGGREGATING ACROSS SOURCES, PATHWAYS, AND/OR ROUTES OF EXPOSURE.	34
6.1 Combining Estimated Exposure Distributions	34
6.2 Limiting Distributional Results	35
7. COMBINING OVER DISJOINT SUBPOPULATIONS	38
8. SUMMARY	41
REFERENCES	42
APPENDIX A	A-1

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Summary of Primary Methods for Characterizing Uncertainty for Exposure Assessments	8
A.1	Selected Percentiles for Estimated Dose Resulting from Equally Likely Input Variable Combinations	A-8
A.2	Selected Percentiles for the Dose Distribution Resulting from the Presumed Model and Input Variables Distributions	A-12
A.3	Selected Percentiles for Estimated Dose Resulting from Alternative Equally Likely Input Variable Combinations. .	A-15
A.4	95 Percent Confidence Interval Estimates for Selected Percentiles of the Distribution of Dose (Y) Based on Model-Predicted Doses	A-19
A.5	95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution (Y) Assuming Uniform and Beta Distributions for IR and ED/600, Respectively. .	A-25
A.6	95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution (Y) Directly from Exposure Data	A-30

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1	Integrated Exposure Assessment	2
2	Risk Assessment/Risk Management Process	4
3	Hypothetical Joint Relative Frequencies for Exposure Duration (ED) and Ingestion Rate (IR)	18
A.1	Distribution of Estimated Doses Resulting from Equally Likely Input Variable Combinations	A-7
A.2	Hypothetical Subjective Estimate of the Distribution of Exposure Duration (ED)	A-9
A.3	Predicted Dose Distribution Resulting from the Presumed Model and Input Variable Distributions	A-11
A.4	Distribution of Estimated Dose Resulting from Alternative Equally Likely Input Variable Combinations	A-14
A.5	Histogram of the Hypothetical Sample of Size 500 for Exposure Duration (ED)	A-17
A.6	Histogram of the Hypothetical Sample of Size 500 for Ingestion Rate (IR)	A-18
A.7	95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution Based on Model- Predicted Doses	A-20
A.8	Confidence Region for the Estimated Beta Distribution Parameters $\hat{\alpha}$ and $\hat{\beta}$	A-23
A.9	95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution Assuming Uniform and Beta Distributions for IR and ED/600, Respectively . . .	A-26
A.10	95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution from Directly Measured Doses	A-31

FOREWORD

The Exposure Assessment Group (EAG) of the U.S. Environmental Protection Agency's (EPA) Office of Research and Development has three main functions: 1) to conduct exposure assessments; 2) to review assessments and related documents; and 3) to develop guidelines for Agency exposure assessments. The activities under each of these three functions are supported by and respond to the needs of the various EPA program offices.

As part of its function to develop guidelines, EAG conducts research projects for the purpose of developing or refining techniques used in exposure assessment. This document presents the results of such a project and serves as a support document to EPA's Interim Guidelines for Exposure Assessment. Specifically, this report describes procedures which can be used to characterize uncertainty in estimated exposure and contains an example to demonstrate the utility of the most complex techniques.

ABSTRACT

Virtually all exposure assessments except those based upon measured exposure levels for a probability sample of population members rely upon a model to predict exposure. The model may be any mathematical function, simple or complex, that estimates the population distribution of exposure or an individual population member's exposure as a function of one or more input variables. Whenever a model that has not been validated is used as the basis for an exposure assessment, the uncertainty associated with the exposure assessment may be substantial. The primary characterization of uncertainty is at least partly qualitative in this case, i.e., it includes a description of the assumptions inherent in the model and their justification. Plausible alternative models should be discussed. Sensitivity of the exposure assessment to model formulation can be investigated by replicating the assessment for plausible alternative models.

When an exposure assessment is based upon directly measured exposure levels for a probability sample of population members, uncertainty can be greatly reduced and described quantitatively. In this case, the primary sources of uncertainty are measurement errors and sampling errors. A thorough quality assurance program should be designed into the study to ensure that the magnitude of measurement errors can be estimated. The effects of all sources of random error should be measured quantitatively by confidence interval estimates of parameters of interest, e.g., percentiles of the exposure distribution. Moreover, the effect of random errors can be reduced by taking a larger sample.

Whenever the latter is not feasible, it is sometimes possible to obtain at least some data for exposure and model input variables. These data should be used to assess goodness of fit of the model and/or presumed distributions of input variables. This substantially reduces the amount of quantitative uncertainty for estimation of the distribution of exposure and is strongly recommended. It is recognized, however, that it may not always be feasible to collect such data.

ACKNOWLEDGMENTS

The author would like to acknowledge that many RTI staff members, not all of whom can be mentioned in this space, contributed to preparation of this report. Several discussion sessions attended by various RTI staff members contributed greatly to the preparation of this report. The author thanks Dr. Martin Massoglia for many helpful suggestions and for contribution of Figures 1 and 2, which he had prepared for other RTI research. Appreciation is extended to Dr. James Chromy, Dr. Douglas Drummond, and Dr. Clark Allen for their suggestions and constructive criticism. Thanks is extended to Mr. Fred Immerman and Mr. George Kircher for their technical assistance in preparation of Appendix A. The author thanks the following peer reviewers for their careful review of the draft version of this document and their many helpful suggestions: Dr. Donald Porcella, Dr. Dominic Ditoro, Dr. John Spengler, Dr. J. Julian Chisolm, Mr. John L. Henshaw; and Dr. Will M. Ollison. The leadership and insightful direction provided by the EPA task manager, Dr. James Falco, is also appreciated. Last, but not least, the author thanks his secretarial staff for their patient preparation of several revisions of this manuscript.

1. INTRODUCTION AND OVERVIEW

An exposure assessment often progresses through several stages of refinement. The purpose of this document is to provide technical guidance for characterization of uncertainty for assessments at various stages of refinement, from assessments based upon limited initial data to those based upon extensive data.

This document is written for the professional who is actively involved in the performance of exposure assessments. It is intended to facilitate use of the latest statistical methods for characterizing uncertainty for exposure assessments.

This introductory chapter begins with a brief section that discusses the nature of an exposure assessment. This section serves as a point of reference for the remainder of this document. It is also intended to make this document essentially self-contained. An overview of the characterization of uncertainty for exposure assessments at various stages of refinement is presented in the final section of this chapter, which also serves as a guide to the remaining chapters.

1.1 WHAT IS AN EXPOSURE ASSESSMENT?

An exposure assessment quantitatively estimates the contact of an affected population with a substance under investigation. Typically, the magnitude, duration, and/or frequency of contact are estimated. The U. S. Environmental Protection Agency (EPA) has prepared a draft handbook [1] to provide guidance pertinent to performing an exposure assessment. In addition, references [2] and [3] are valuable sources of information with regard to the methodology of exposure assessments.

An integrated exposure assessment quantifies the contact of population members with the substance under investigation via all routes of exposure (inhalation, ingestion, and dermal) and all pathways from sources to exposed individuals (e.g., from manufacturing plants to plant workers and/or consumers of manufactured products). As shown in Figure 1, an integrated exposure assessment also estimates the size of population affected by each exposure route. Since an integrated exposure assessment is often needed to enable regulatory decisions, characterization of uncertainty is addressed in this context.

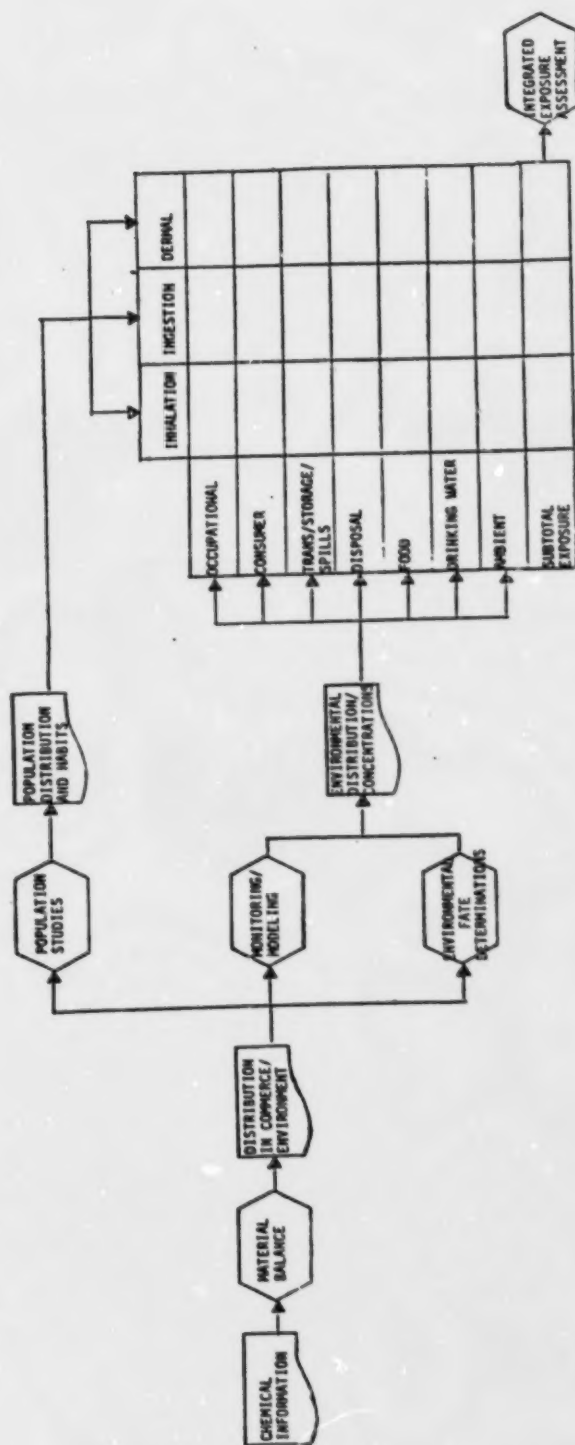


Figure 1. Integrated Exposure Assessment

The overall uncertainty associated with both the exposure assessment and the dose-response relationships must be addressed by risk assessments. Characterization of uncertainty for exposure assessments is addressed by this document; other comparable documents address characterization of uncertainty for risk assessments per se. The primary purpose of an exposure assessment is to enable risk assessment and possible selection of a regulatory option. Regulatory decisions must be capable of withstanding close scrutiny (perhaps even in a court of law). Therefore, the characterization of uncertainty for an exposure assessment should be very thorough and based upon state-of-the-art quantitative methods to the extent possible.

The type of exposure (internal dose, external dose, etc.) investigated by an integrated exposure assessment depends upon the input needed to enable risk assessment (see Figure 2). In most cases an exposure assessment should ideally quantify the dose of the substance under investigation that is absorbed by individual body organs in order to enable the most reliable estimates of health effects. Presently it is practical to quantify absorbed organ-specific doses only in certain circumstances--e.g., when the affected organ is the skin (dermal route) or when exposure is due to radiation. Moreover, the absorbed whole-body dose often cannot be quantified. Instead, indicators of the absorbed dose are often used by necessity in exposure assessments. These indicators of absorbed doses may be either environmental exposure or body burden ([3], [4]). Environmental exposure refers to the levels of the substance under investigation contacted through the air, water, food, soil, etc. Body burden refers to levels of the substance in body fluids or tissues, e.g., blood, urine, breath, fatty tissues, hair, nails, etc. In the remainder of this document, the term "exposure" will usually be used in a generic sense to represent either the absorbed internal dose or an indicator of the absorbed dose because both have been used as the basis for exposure assessments.

An integrated exposure assessment generally partitions the affected population into subpopulations such that all members of a subpopulation are affected by the same sources of exposure. For each source, there can be one or more pathways by which the substance under study travels from the source to the affected population. For each source and pathway, a population member can be exposed via one or more of three absorption routes: inhalation, ingestion, and dermal. An integrated exposure assessment often begins by estimating exposure for a subpopulation via a specific combination of source, pathway,

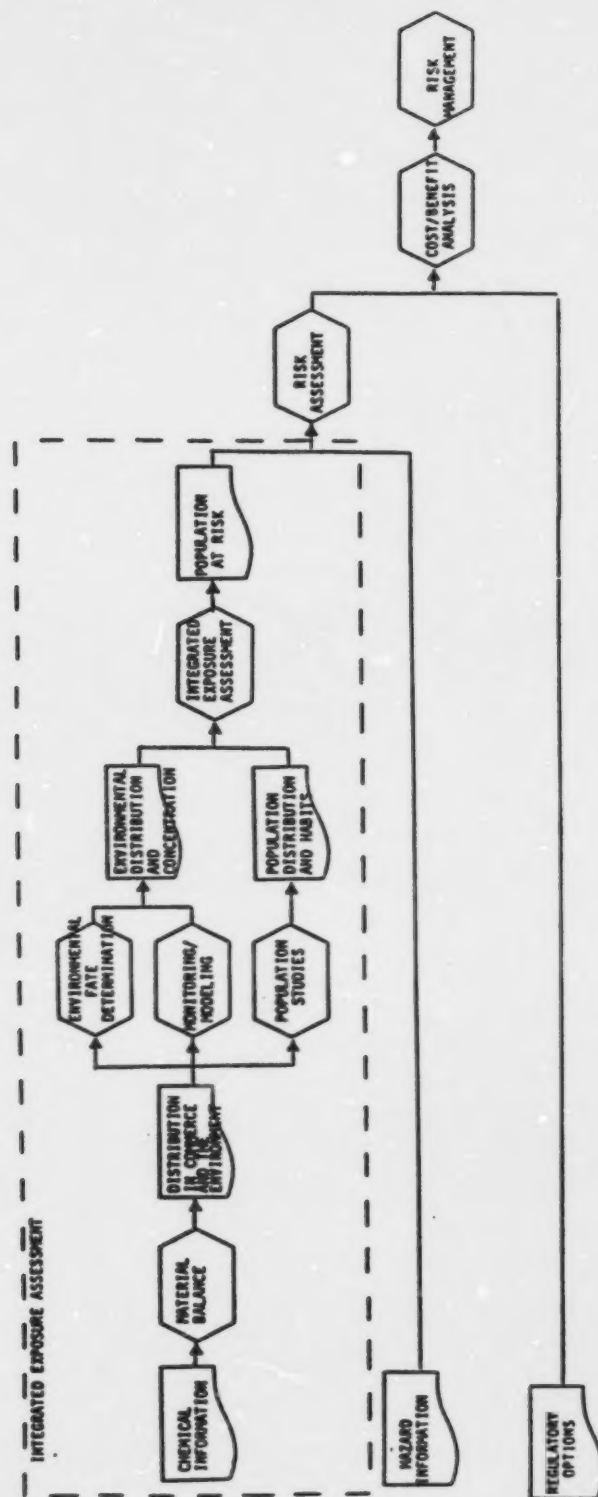


Figure 2. Risk Assessment/Risk Management Process

and route. Exposure is then aggregated across pathways and sources to estimate exposure via a particular route for subpopulation members. Exposure may also be aggregated across routes if this aggregate is relevant to estimation of risk. Finally, the sizes of the subpopulations should also be estimated.

An exposure assessment addresses the exposures experienced by the subjects (human or nonhuman) of a specified population. The most complete characterization of these exposures is the population distribution of individual exposures. If the population distribution of exposure was known, then all summary parameters of the distribution would also be known, such as the following:

1. Minimum of all possible exposures.
2. Maximum of all possible exposures.
3. Mean, median, and mode of the population exposures levels.
4. Standard deviation of the exposure distribution.
5. Proportion of subjects whose exposure exceeds any specified exposure level.
6. Percentiles of the exposure distribution.

Since knowledge of the distribution of exposures implies knowledge of all the above parameters that are so useful for evaluation of risk, estimation of this distribution is a primary goal of exposure assessments. However, some exposure assessments may be limited by the available data to estimating summary parameters of the exposure distribution, e.g., the mean exposure or certain percentiles of the exposure distribution.

1.2 CHARACTERIZATION OF UNCERTAINTY

Reference [5] addresses characterization of uncertainty for a source severity index to be used as an aid in regulatory decision making. Interval estimation for the mean value of the severity index based on "error propagation formulas" is discussed. Assuming that the index is a function of random variables for which direct confidence interval estimates are available, a first-order Taylor series approximation of the function is recommended for computation of an approximate confidence interval estimate for the mean value of the index. A Taylor series approximation is also recommended for derivation of bounds on the systematic error, or bias, in estimation of the mean severity index. Tables of error propagation formulas for random errors and

systematic errors are provided. Reference [1] recommends these methods for characterizing uncertainty for exposure assessments. However, these methods have limited applicability for the following reasons: (1) The primary purpose of an exposure assessment is usually to estimate the distribution of exposure, not a mean severity index, and (2) direct confidence interval estimation of the mean index would be preferable to using "error propagation formulas." The present document will concentrate on estimation of the distribution of exposures and on characterization of the uncertainty associated with this estimated distribution.

A review of methodology for characterizing uncertainty is provided by reference [6]. This article discusses uncertainty for a predicted output variable Y when the output can be expressed as a deterministic function of several random variables. Since it is the uncertainty of the predicted output that is of interest in this article, the probability distribution of the output variable and the variance of this distribution are advocated as primary characterizations of the uncertainty associated with the predicted output. Within this context, the estimated distribution of exposures could be regarded as a characterization of the uncertainty with regard to the prediction of an individual population member's exposure level, Y. However, estimation of the population distribution of exposures is the primary goal of an exposure assessment, and it is the uncertainty with regard to this estimated distribution of exposures that must be addressed. Thus, reference [6] is a valuable reference with regard to characterization of uncertainty, but the present document will rely mainly upon somewhat different methods that more directly address characterizing uncertainty for an estimated distribution of exposures.

The present document is not intended to be a general review of methodology for characterizing uncertainty. Instead, this document focuses on those methods for characterizing uncertainty that are most appropriate for exposure assessments. The determination of which methods are appropriate for characterizing the uncertainty of an estimate depends upon the underlying parameters being estimated, the type and extent of data available, and the estimation procedures utilized. Therefore, this document is largely organized according to the types of data and estimation procedures that may reasonably form the basis for an exposure assessment. As a result, a great deal of this document

actually addresses estimation procedures for exposure assessments. The methodology considered most appropriate for characterizing the uncertainty associated with an exposure assessment is then discussed with regard to each estimation procedure. For example, when the population distribution of exposure is being estimated, characterization of uncertainty addresses the possible differences between the estimated distribution of exposures and the true population distribution of exposures.

An exposure assessment is often based on one or more exposure scenarios and associated mathematical models. An exposure scenario ideally considers all potential sources of the substance that could come in contact with population subjects. Exposure prediction models based upon transportation and fate of the substance would then be used to describe the pathways from sources to contact with population members. These models may be very complex, using data on levels of the substance at sources and/or ambient monitoring sites and the geographic distribution of the target population to estimate the population distribution of exposures ([1], [7], [8], [9]). Alternatively, these models can be relatively simple models, such as multiplicative factors, that predict an individual population member's personal exposure as a function of several input variables (see, e.g., references [3] and [10], and Appendix A).

Table 1 provides an overview of the primary methods recommended for characterizing uncertainty for exposure assessments. Both qualitative and quantitative methods for characterizing uncertainty are presented for each of the following five types of data:

1. Measured personal exposure levels for a sample of population members.
2. Measured model input variables for a sample of population members, given a model that predicts personal exposure as a function of input variables.
3. Estimated joint distribution of model input variables, given a model that predicts personal exposure as a function of the input variables.
4. Limited data for model input variables, given a model that predicts personal exposure as a function of the input variables.
5. Data on levels of a substance at fixed sites (possibly including sources) and geographic distribution of population members, given a model that predicts the population exposure distribution from such input data.

Table 1. SUMMARY OF PRIMARY METHODS FOR CHARACTERIZING UNCERTAINTY FOR EXPOSURE ASSESSMENTS

Type of data utilized	Primary methods for characterizing uncertainty ^a	
	Qualitative methods	Quantitative methods
Measured personal exposure for a sample of population members.	1. Limitations of the survey design and measurement techniques.	1. Confidence interval estimates for percentiles of the exposure distribution and/or other parameters of the exposure distribution. 2. Estimation of measurement accuracy and precision, and the consequences of potential measurement bias.
Measured model input variables for a sample of population members, given a model that predicts personal exposure as a function of the input variables.	1. Limitations of the survey design and measurement techniques. 2. Validity of the personal exposure model.	1. Confidence interval estimates for percentiles of the exposure distribution and/or other parameters of the exposure distribution. 2. Goodness of fit of the personal exposure model if measured exposures are also available for any of the sample members. 3. Estimation of measurement accuracy and precision, and the consequences of potential measurement bias. 4. Estimation of the sensitivity of parameter estimates to plausible alternative formulations of the model.
Estimated joint distribution of model input variables, given a model that predicts personal exposure as a function of the input variables.	1. Validity of the personal exposure model. 2. Limitations of the data or other basis for the input variable distributions.	1. Goodness of fit of the estimated joint distribution of model input variables if data are available for the input variables for a sample of population members. 2. Estimation of the sensitivity of parameter estimates to plausible alternative formulations of the model. 3. Goodness of fit of the personal exposure model if data are available for both input variables and personal exposure for a sample of population members.
Limited data for model input variables, given a model that predicts personal exposure as a function of the input variables.	1. Validity of the personal exposure model. 2. Limitations of the data.	1. Estimation of the sensitivity of parameter estimates to plausible ranges of the input variables values and/or plausible alternative models.

Table 1. (continued)

Type of Data Utilized	Primary Methods for Characterizing Uncertainty ^a	
	Qualitative Methods	Quantitative Methods
Data on levels of a substance at fixed sites (possibly including sources) and geographic distribution of population members, given a model that predicts the population exposure distribution from such input data.	1. Validity of the population exposure model.	1. Estimation of the sensitivity of parameter estimates to plausible alternative formulations of the model.
	2. Limitations of the input variable data.	2. Validation of the model, confidence interval estimation, and estimation of measurement accuracy and precision are all desirable, but such quantitative characterization of uncertainty may not be possible with this type of data.

^aIt is assumed that the purpose of the exposure assessment is to estimate the distribution of exposure and/or summary parameters thereof (e.g., mean, 5th and 95th percentiles, maximum possible exposure, proportion of population with exposure exceeding a specified level, etc.).

In each case, it is assumed that the purpose of the exposure assessment is to estimate the distribution of exposure and/or summary parameters thereof (e.g., mean, 5th and 95th percentiles, maximum possible exposure, proportion of population with exposure exceeding a specified level, etc.).

The initial exposure assessment for a substance is often based upon a very limited amount of data. Under such conditions, the primary characterization of uncertainty may be qualitative. Exposure assessments based upon limited data are discussed in Chapter 2.

Chapter 3 then discusses assessments based upon the estimated joint probability distribution of model input variables, given a model that predicts personal exposure as a function of the input variables. Depending upon the extent of data available to validate the model as well as the input variable distributions, this type of assessment has the potential for reducing uncertainty relative to the methods discussed in Chapter 2. It also has the potential for a more quantitative characterization of the uncertainties.

The uncertainties associated with an exposure assessment can potentially be reduced further by basing the assessment on measured values of model input variables for a sample of population members, given a model that predicts personal exposure as a function of the input variables, as discussed in Chapter 4. Finally, Chapter 5 discusses exposure assessments based on measured exposure levels for a sample of population members, which has the potential for minimizing the uncertainties associated with an assessment.

When the initial exposure assessment for a substance is performed, only limited data relating to exposure will usually be available. As more resources are devoted to an exposure assessment, a more refined assessment may be generated. There may be many stages of refinement for an exposure assessment. The degree of refinement ultimately needed for an exposure assessment depends upon the accuracy needed to enable risk management decisions, as well as the cost of obtaining adequate data, the time available for data acquisition, etc.

When the exposure levels experienced by individual population subjects are measured, the measured exposures generally reflect the total exposure across all sources and pathways, and possibly across all routes as well. However, when the modeling approach is employed for an exposure assessment, a model often predicts an individual population member's exposure due to a specific combination of source, pathway, and route. Methods for combining

estimated exposure distributions over sources, pathways, and routes are discussed in Chapter 6. Finally, Chapter 7 discusses combining the estimated exposure distributions over disjoint subpopulations to produce an estimate of the exposure distribution for the total population. Methods for characterizing the uncertainties associated with these estimated exposure distributions are also discussed in Chapters 6 and 7.

A hypothetical example of a subpopulation exposure assessment that progresses through several stages of refinement is presented in Appendix A. The sections of the example in Appendix A correspond directly to the sections of the first five chapters of this report and are numbered accordingly (e.g., Section 4.1 of Appendix A corresponds directly to Section 4.1 of Chapter 4).

2. ASSESSMENTS BASED UPON LIMITED DATA

The initial exposure assessment for a substance may be based upon limited data for directly measured personal exposures, concentrations in the environment, and/or variables that can be used to predict personal exposure. These data may be either extant data or data produced by a small-scale study. Characterization of uncertainty for such an initial exposure assessment depends mainly upon whether the data are direct measurements of exposure or measurements of variables that can be used to predict exposure. The purpose of this chapter is to discuss characterization of uncertainty in each of these situations when data are limited. Characterization of uncertainty for an assessment based upon directly measured exposures for a small sample of individuals is discussed first. Characterization of uncertainty is then discussed for estimates of the distribution of exposure based upon limited data for model input variables. Both the population exposure models that directly estimate the population distribution of exposure and the personal exposure models that estimate the exposure experienced by a sample of population members are addressed.

2.1 LIMITED DATA FOR DIRECTLY MEASURED EXPOSURES

Historically, the discovery of a new environmental pollutant often begins when a physician recognizes a clustering of cases with common clinical features that are unique and unexpected for the community in which the clustering has occurred. If the physician reports this situation to the appropriate public health authorities, an initial study on a small sample of unusual cases may be undertaken. Alternatively, if the physician reports the situation in the medical literature, a followup study by public health authorities may result. Under these circumstances, the initial assessment is likely to be based on limited extant data, i.e., data for a sample of the unusual cases and standard baseline laboratory values. If a potential health problem was discovered, a more refined assessment based upon a probability sample¹ of population members would almost always be needed to permit defensible risk management decisions.

¹A probability sample is a sample for which every population member has a chance (i.e., positive probability) of inclusion in the sample. The simplest example is a simple random sample, in which every population member has the same positive probability of inclusion in the sample. See reference [11] for an introduction to probability sampling methods.

If an assessment is based upon a limited database of measured personal exposures, a primary aspect of the uncertainty characterization is whether or not the data result from a probability sample. Characterization of uncertainty for an assessment based upon a valid sample of directly measured exposures is discussed in Chapter 5. Statistics calculated from a small probability sample generally have relatively large standard errors, i.e., a relatively high level of uncertainty. However, a small probability sample provides the basis for more rigorous quantitative characterization of uncertainty than would be possible for a small non-probability sample.

If the measured personal exposure levels for a small sample of population members were not so low that health risks were known to be negligible, a more refined exposure assessment might be pursued. Two ways of refining the exposure assessment would be as follows:

1. Measure exposure levels for a large enough sample to make inferences with a predetermined level of precision using the methods discussed in Chapter 5.
2. Develop an exposure scenario and associated exposure prediction model to be used as the basis for an assessment as discussed from Section 2.2 through Chapter 4.

Since the exposure variable needed for predicting risk is often not amenable to direct observation, the latter model-based approach is often pursued first, if not exclusively. Hence, the model-based approach to refining an assessment is discussed first in this document.

2.2 LIMITED DATA FOR MODEL INPUT VARIABLES

2.2.1 Population Exposure Models

An initial exposure assessment is often based upon extant data or other easily obtainable data regarding ambient concentrations and/or emission levels for known sources of the substance in question. Such data may then be combined with data on the geographic distribution of population members to estimate the population distribution of exposures using an elaborate population exposure model ([1], [7], [8], [9], [12], [13], [14]). Such exposure assessments tend to be subject to a high level of uncertainty. However, the characterization of uncertainty for these assessments is often primarily qualitative due to the nature of the underlying data and the estimation methods.

Characterization of uncertainty for assessments based upon population exposure models consists primarily of a description of the limitations of the population exposure model and limitations of the data for the model input

variables (see especially reference [12]). In addition, sensitivity of the predicted exposure distribution to plausible changes in the model and/or input variable values can be quantified (see references [13] and [14]). Additional quantitative characterization of uncertainty may not be possible, but validation of the exposure model, confidence interval estimates of relevant parameters of the exposure distribution, and estimation of the accuracy and precision of input variable measurements should all be pursued whenever possible.

2.2.2 Personal Exposure Models

An initial exposure assessment may be based upon limited data (e.g., estimated ranges) for the input variables of a model that predicts personal exposures. The personal exposure model would be based upon an exposure scenario that considers all potential sources of the substance that could result in contact with population members. Knowledge of transportation and fate of the substance would be used to model the pathways from sources to contact with population subjects. These models can be very simple, such as models involving only multiplicative factors (see, e.g., references [3] and [10], and Appendix A). A personal exposure model is any model, simple or complex, that expresses an individual population member's exposure as a function of one or more input variables.

If the available data are only sufficient to support estimates of the ranges of the input variables, the exposure assessment might be limited to a sensitivity analysis. The purpose of the sensitivity analysis would be to identify influential model input variables and develop bounds on the distribution of exposure. A sensitivity analysis would estimate the range of exposures that would result as individual model input variables were varied from their minimum to their maximum possible values with the other input variables held at fixed values, e.g., their mid-ranges. In addition, the overall minimum and maximum possible exposures would usually be estimated.

The uncertainty for an exposure assessment of this type would be characterized by presenting the ranges of exposure generated by the sensitivity analysis, by describing the limitations of the data used to estimate plausible ranges of model input variables, and by discussing justification for the model. Justification of the model should include a description of the exposure scenario, choice of model input variables, and the functional form of the model. Sensitivity to the model formulation should also be investigated by replicating the sensitivity analysis for plausible alternative models.

If the maximum possible exposure estimated by the sensitivity analysis presented no potentially significant health risk, there might be no need to refine the assessment. If both the minimum and maximum exposures presented a potentially significant health risk, it would be known that the exposure scenario represented a significant health problem without refining the assessment. When the minimum does not present a potentially significant health risk and the maximum does, then greater importance is placed on choosing a summary parameter of the exposure distribution (e.g., the mean or a specific percentile) as the basis for a regulatory decision. A more in-depth exposure assessment enabling estimation of the distribution of exposures permits selection of any summary parameter (minimum, maximum, mean, or a percentile, etc.) as the basis for a regulatory decision, and permits the decision maker to select that summary parameter after the fact (i.e., after the exposure distribution has been estimated).

The sensitivity analysis can be enhanced by computing the predicted exposures that result from all possible input variable combinations. If each input variable has only a finite set of possible values, the set of all possible combinations of the input variables can be formed, and the predicted exposure can be computed for each combination. The input variable combinations resulting in exposures that constitute potential health problems can thereby be identified. These exposure predictions can be used to form a probability distribution of exposures, which will be referred to as the exposure sensitivity distribution, by computing the relative frequency of occurrence of each exposure level or interval of exposures. The resulting exposure sensitivity distribution is equivalent to the population distribution of exposures only if all combinations of the input variable values are equally likely to occur in the target population. The exposure sensitivity distribution can also be derived for personal exposure models based upon continuous input variables with finite ranges by representing these variables by equally-spaced points. At the limit as the equal spaces become infinitely small and the number of points becomes infinitely large, the procedure for computing the exposure sensitivity distribution is equivalent to estimating the probability distribution of exposures that results from statistically independent continuous input variables with Uniform distributions on the estimated ranges. Thus, the exposure sensitivity distribution can be derived analytically or by Monte Carlo simulation based upon independent Uniform input variable distributions.

The Monte Carlo method consists of randomly generating input variable values and using these to compute corresponding exposure levels, generating the exposure sensitivity distribution via replications.

Statistics based upon the exposure sensitivity distribution should be interpreted in terms of all possible input variable combinations. For example, the 95th percentile of the exposure sensitivity distribution would be the exposure level exceeded by only 5 percent of the exposures resulting from all possible combinations of input variable values. Although this distribution of exposures is not an estimate of the population distribution (unless all possible combinations of the input variables actually are equally likely), it has more utility for regulatory decision making than a standard sensitivity analysis. For example, if 99 percent of the computed exposures are below the level that represents a significant health concern, the proportion of the population at risk is almost surely small (albeit possibly smaller or larger than 1 percent).

Characterization of uncertainty for the enhanced sensitivity analysis would include a discussion of limitations of the data and justification for the model as discussed above for the usual sensitivity analysis. Sensitivity to model formulation should also be investigated by estimating the distributions of exposure that result from using the same Uniform input variable distributions with plausible alternative models and comparing the estimated percentiles of the induced exposure distributions.

3. ASSESSMENTS BASED UPON ESTIMATION OF INPUT VARIABLE DISTRIBUTIONS

If a model has been formulated that expresses an individual's personal exposure as a function of one or more input variables, the population distribution of exposure can be estimated analytically or by Monte Carlo simulation from an estimate of the joint distribution of the model input variables.² A distribution of exposure derived in this way is not equivalent to one based upon measured exposures from a personal monitoring study. The derived distribution merely reflects the assumed personal exposure model and the probability distributions used to represent the model input variables. All such assumptions should be explicitly stated and justified. Ideally, model input variables should be represented by empirically validated probability distributions. However, the joint distribution of model input variables can be estimated from discussions with subject-matter experts (e.g., via univariate histograms when the input variables may be treated as statistically independent). The estimated population distribution of exposures will be equivalent to the exposure sensitivity distribution discussed in Section 2.2.2 only for the case when all the input variable distributions actually are independent Uniform distributions. Hence, the purpose of this chapter is to characterize uncertainty for exposure assessments based upon personal exposure models that can be used to estimate the population distribution of exposures from information concerning input variable distributions.

3.1 ESTIMATION BASED ON INPUT VARIABLE DISTRIBUTIONS

When the data for model input variables are limited, discussions with subject-matter experts can be used to formulate an estimate of the joint distribution of the input variables. Such discussions can be used to investigate independence of the input variables and determine the approximate location and shape of the input variable distributions. Qualitative knowledge of the input variable distributions can then be represented quantitatively by input variable probability distributions. The input variable distributions should be validated quantitatively using goodness-of-fit tests whenever sufficient data are available (see, e.g., Chapter 8 of reference [15], or Section 4.5 and Chapter 6 of reference [16]).

²Equivalently, one can estimate the marginal distributions, i.e., the distribution of each input variable ignoring the others, if the input variables can be treated as independent.

Discussions with subject-matter experts for the purpose of estimating the joint distribution of model input variables should first determine whether or not the model input variables can be treated as statistically independent for the population of interest. If model input variables are assumed to be independent, the theoretical and/or empirical justification of the assumed independence should be discussed. If two model input variables cannot be treated as independent, their joint probability distribution can be estimated by determining the relative frequency of occurrence for areas in the cross product space defined by the two variables. As a hypothetical example, the joint probability distribution of ingestion rate (IR) and exposure duration (ED) could be estimated for the model in Appendix A by the relative frequencies of occurrence shown in Figure 3. (These relative frequencies are hypothetical for the purpose of illustration only; these variables are treated as independent variables in Appendix A.) In addition, the marginal distribution of each of the correlated variables, i.e., the distribution of each variable ignoring the others, should be derived from the joint distribution and reconciled with available knowledge of the marginal distributions. Of course, if the input variables are not highly correlated or if the model is not sensitive to the assumption of independence, it may be sufficient to treat the variables as independent and to acknowledge minor departures from the independence assumption as a source of uncertainty.

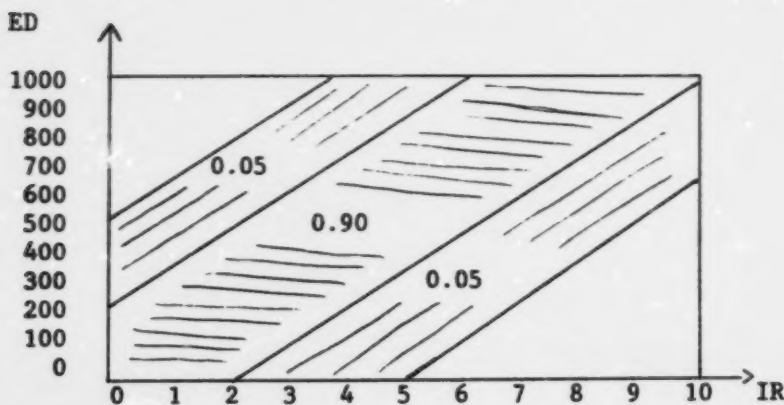


Figure 3. Hypothetical Joint Relative Frequencies for Exposure Duration (ED) and Ingestion Rate (IR)

The input variables of an exposure model that are statistically independent of all other input variables can be considered individually by subject-matter experts, one variable at a time. The expert that is consulted with regard to establishing a probability distribution for a factor in an exposure model need not be conversant in statistics. A probability distribution can be estimated from such information as the following: general shape of the distribution, smallest possible value, largest possible value, most likely value (mode), average value (mean), the median value such that half the values of the variable are larger and half are smaller, and/or percentiles of the probability distribution (e.g., the 5th and 95th percentiles). The subject-matter expert need not be able to specify all of these statistics in order to estimate the distribution. In some cases, the subject-matter expert may be able to provide a histogram to be used as the probability distribution of a model input variable.

Such subjective probability distributions should be confirmed in discussions with other subject-matter experts. The Delphi technique can be used to obtain the most reliable consensus of opinion from a group of experts by using a series of questionnaires interspersed with controlled feedback of group opinion [17]. If such discussions do not yield a consensus of opinion, sensitivity of the model to the alternative input variable distributions suggested by experts should be investigated. Alternatively, direct measurements of the model input variables and/or exposure itself may be used to estimate the population distribution of exposures as discussed in Chapters 4 and 5.

Discussions with subject-matter experts, empirical data, and/or theoretical arguments may suggest that a particular family of probability distributions is appropriate for representing a model input variable. For input variables with a finite range of possible values, distributions such as the Uniform, Beta, Truncated Normal, or Truncated Lognormal may be supported. The parameters of such distributions can usually be estimated from a measure of central tendency (e.g., mean, median, or mode), a measure of variability (e.g., standard deviation or a percentile), and an estimate of the finite range of possible values. Reference [18] provides guidance with regard to estimation of parameters of continuous probability distributions.

The following considerations may be useful when trying to determine an appropriate probability distribution to represent a model input variable:

1. A histogram distribution may be a very useful initial estimate for the distribution of a model input variable.

2. If a model input variable has a finite range of possible values and if a smooth probability function is desired, a Beta distribution (scaled to the desired range) may be appropriate. The Beta probability distribution is very flexible; its probability function can assume a variety of shapes by varying its parameters.
3. If an input variable only assumes values greater than some minimum value (e.g., is positive-valued) then a Gamma, Weibull, or Lognormal distribution may be appropriate. The Gamma distribution is more flexible, and its probability function can assume a variety of shapes by varying its parameters.
4. If an input variable has an unrestricted range of possible values, a Normal, Student's t, or Logistic distribution may be an appropriate probability distribution. These distributions differ mainly in their tails. In addition, truncated versions of these variables may be used in situations 2 and 3 described above.

The example in Appendix A uses Monte Carlo simulation to predict the distribution of exposure from presumed independent distributions for exposure duration (ED) and ingestion rate (IR), the model input variables. For a population of N members, each person would have his own individual value for each of these variables. The actual joint distribution of the input variables would then be the distribution of the N paired values of ED and IR experienced by the individual population members. If the correlation of these variables was not negligible, the joint distribution of these variables would have to be estimated (e.g., using joint relative frequencies such as shown in Figure 3) in order to produce a reliable estimate of the population distribution of exposure using the personal exposure model.

When qualitative knowledge of the input variable distributions is used to estimate the population distribution of exposure, uncertainty is characterized by discussing justification for the presumed model and input variable distributions. The choice of model input variables and the functional form of the model should be discussed. Empirical data and/or theoretical arguments that support the model should be presented. Use of a model and input variable distributions that have been empirically validated is recommended in order to reduce uncertainty. Alternative models and/or alternative input variable distributions should also be discussed. Sensitivity to these alternatives can be investigated by estimating the distributions that result from plausible alternatives and comparing the percentiles of the induced exposure distributions. All available data, even if the data are limited, should be used to validate the presumed input variable distributions and the predicted distribution of exposure.

3.2 LIMITING DISTRIBUTIONAL RESULTS

Given a mathematical model that expresses personal exposure as a function of several input variables, the distribution of exposure can be derived analytically or by using Monte Carlo simulation as described above. The distribution of exposure is then induced by the distributions of the model input variables. However, as the number of input variables becomes large, it can be shown that the induced exposure distribution approaches a specific limiting distribution for certain classes of models. The true exposure distribution for a model based upon several input variables may then be approximated by the limit of a sequence of conceptual models with more and more input variables. For these classes of models, the means, variances, and covariances are the most important characteristics of the input variable distributions because they are sufficient to determine the limiting distribution. Other shape characteristics of the input variable distributions then become of lesser importance for models with several input variables.

Limiting distributional results can easily be derived for models that can be represented as a special case of the following model (such as the models discussed in references [3] and [10], and Appendix A),

$$Y = \prod_{i=1}^n X_i / \prod_{i=n+1}^d X_i, \quad (1)$$

where all input variables, X_1 through X_d , are independent positive-valued random variables (which includes positive constants as a special case). Since all input variables are positive, this model can equivalently be expressed as

$$\ln Y = \sum_{i=1}^n \ln X_i - \sum_{i=n+1}^d \ln X_i. \quad (2)$$

It then follows that if the independent random variables $\ln X_1$ through $\ln X_d$ are such that a central limit type of result applies (i.e., there is no one input variable ($\ln X_i$) whose variance is large relative to the variance of $\ln Y$), then the limiting distribution of the exposure, Y , is lognormal (see, e.g., Section 14-1.1 of reference [19]). Moreover, if the random variables X_1 through X_d have comparable variances, then the convergence to lognormal may be rapid, i.e., the product of two or three factors could yield an exposure distribution that is approximately lognormally distributed. It should be noted, however, that this convergence to lognormality applies to the population distribution of exposure only so long as the model, Equation (1), accu-

rately reflects the true state of nature and so long as there is no single input variable whose variance dominates the total variance. Data for exposure and/or model input variables should be used to validate the model whenever possible.

Since the mean and variance are sufficient statistics for the lognormal distribution, estimation of these parameters will completely specify the limiting lognormal distribution when the exposure model is of the form given by Equation (1). In fact, any measure of central tendency (e.g., mean, median, or mode) together with any measure of dispersion (e.g., standard deviation or a percentile) can be used to estimate the parameters of the limiting lognormal distribution. When directly measured exposure data are available for individual population subjects, these data may be used to directly estimate the parameters of the limiting lognormal distribution. It is also possible to use the means, variances, and covariances of the model input variables to estimate the parameters of the limiting lognormal distribution of exposure without formulating specific probability distributions for the model input variables, as shown below.

A consistent estimate of the mean of Y is the ratio

$$\hat{E}_c(Y) = \frac{\prod_{i=1}^n \hat{E}(X_i)}{\prod_{i=n+1}^d \hat{E}(X_i)}, \quad (3)$$

where $\hat{E}$ denotes an unbiased estimate of the mean (an estimate for which the expected value over all possible samples is the population mean) and $\hat{E}_c$ is a consistent estimate of the mean (an estimate that converges to the population mean as the sample size increases). When the input variables can be treated as independent, the numerator and denominator estimates are unbiased, but the ratio is a biased estimate of Y. When the input variables cannot be treated as independent, standard practice would be to continue to use Equation (3) anyway, which still yields a consistent estimate of Y.

A consistent estimate of the variance ($\hat{V}_c$) of the exposure distribution can be obtained by using a first-order Taylor series approximation. This variance approximation may be expressed in general for a function of n random variables as follows:

$$\begin{aligned} \hat{V}_c[F(X_1, \dots, X_n)] &= \sum_{i=1}^n \left[\frac{\partial F}{\partial X_i}(\bar{X}) \right]^2 \hat{V}(X_i) \\ &+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[\frac{\partial F}{\partial X_i}(\bar{X}) \right] \left[\frac{\partial F}{\partial X_j}(\bar{X}) \right] \hat{Cov}(X_i, X_j), \end{aligned} \quad (4)$$

where ∂ is the partial derivative operator; $\bar{X}$ is the mean of $X_1, \dots, X_n$; $\hat{V}$ and Cov denote unbiased estimates of variance and covariance; and V_c is a consistent estimate of variance (see, e.g., reference [20] and Section 10.6 of reference [21]). This result can be applied directly to the model given by Equation (1) because this model is easily differentiable with respect to the input variables, X_i . Moreover, whenever the model input variables are all independent, as is often presumed to be true, the second term in Equation (4) becomes zero.

It should be noted that the Taylor series variance approximation, Equation (4), is the basis for the "propagation of error formulas" found in references [1] and [5]. The application here is very different. References [1] and [5] use the Taylor series to approximate the variance of the mean of model predictions as a characterization of the uncertainty with regard to the population mean. The present application is for estimation of the population variance of the exposure distribution when the basic form of the exposure distribution, e.g., lognormal, has already been established. Thus, Equation (4) is being used to facilitate estimation of the population distribution of exposure. The uncertainty with regard to this estimated distribution is the uncertainty that must be addressed, not that for the estimated mean of the distribution.

Consistent estimates of the mean and variance of Y from Equations (3) and (4) can be used to fit the limiting lognormal distribution of Y directly. No specific distributional form need be assumed for the model input variables, X_i . As noted earlier, any estimate of the central tendency of Y , together with any estimate of variability or a percentile of the distribution of Y , can also be used to directly estimate the limiting distribution of Y . Moreover, whenever the estimate of risk simply involves multiplication of the exposure by one or more positive factors, these results extend directly to estimation of the probability distribution for risk.

The lognormal distribution for exposure Y is attained for the exposure models given by Equation (1) as the number of model input variables becomes infinitely large. In practice, the number of factors in a model will always be finite. The importance of the limiting lognormal result is that the predicted exposure distribution will be approximately lognormal when there are several input variables, none of which dominates the variability of exposures. Thus, when the model has a sufficient number of factors, the means, variances,

and covariances of the input variable distributions are typically sufficient to determine the limiting distribution. Other shape characteristics of the input variable distributions then become less important.

Although the limiting distribution could be estimated directly for a model with several factors, research into the rate of convergence to the limiting distribution would be needed for proper interpretation. Empirical research would be needed to determine when a model had a sufficient number of factors to use the limiting distribution as an adequate approximation to the true distribution. The results would likely depend upon the type of model and input variables.

If Equations (3) and (4) are used to estimate the limiting lognormal exposure distribution for a model of the type given by Equation (1), adequacy of the approximation of the true exposure distribution by the limiting distribution should be discussed. The results of an empirical study on convergence to the limiting distribution might be cited. If data on measured exposures for a sample of population members were available, these data should be used to test the adequacy of the limiting distribution. Characterization of uncertainty with regard to the estimated exposure distribution would also need to address uncertainty with regard to the estimated mean and variance given by Equations (3) and (4). If possible, confidence interval estimates of the mean and variance should be derived. Sensitivity of the predicted exposure distribution to plausible ranges of values for these parameters should then be investigated. Justification of the presumed exposure model should be discussed, as well as limitations of the data used to estimate means and variances or to validate the model or limiting distribution.

4. ASSESSMENTS BASED UPON DATA FOR MODEL INPUT VARIABLES

The exposure assessment based upon an estimate of the joint probability distribution for the input variables of a personal exposure model can be refined by collecting sample survey data for all model input variables. The population distribution of exposure can then be estimated by computing the model-based exposure for each sample member. These exposure values can be used to directly compute confidence interval estimates for percentiles of the exposure distribution. If specific families of probability distributions are accepted as representing the joint distribution of the model input variables, sample survey data can be used to compute joint confidence interval estimates for the parameters of the input variable distributions. These interval estimates for distribution parameters can then be used to compute confidence interval estimates for percentiles of the exposure distribution. Using either method of estimation, interval estimates for percentiles of the exposure distribution are a useful quantitative characterization of uncertainty for the exposure assessment.

The purpose of this chapter is to discuss characterization of uncertainty for exposure assessments based upon model input data collected for a sample of population members. Computation of confidence interval estimates for percentiles of the exposure distribution directly from model-based exposures for the sample subjects is considered first. Derivation of equivalent confidence intervals from confidence intervals on parameters of the input variable distributions is then considered in Section 4.2.

4.1 INTERVAL ESTIMATES OF EXPOSURE PERCENTILES: INPUT VARIABLE DISTRIBUTIONS NOT KNOWN

For many personal exposure models, it may be more feasible and less costly to collect data on the model input variables, rather than on the exposure itself. The population may be thought of as a set of N subjects, each of which experiences specific values of the model input variables and resulting exposure. If the input variable values have been recorded for a sample of population subjects (see, e.g., the hypothetical data in Attachment A.1 of Appendix A), then a model can be used to compute the predicted exposure for each sample member. Treating these computed exposures as if they were measured exposures, the methods described in Section 10.18 of [22] and Section 12.9 of [23] can be used to compute confidence interval estimates for percentiles of the exposure distribution.

Interval estimates for percentiles of the exposure distribution are a useful quantitative characterization of uncertainty for the exposure assessment. However, interval estimates computed as discussed in this section are based upon several assumptions, primarily the presumed model for personal exposures. Hence, characterization of uncertainty for the exposure assessment requires justification of the model and other assumptions. The empirical and/or theoretical basis for the model should be discussed. Use of a validated model always decreases the overall uncertainty. Design of the sample survey that is used to produce the data should also be discussed, as well as accuracy and precision of the measurement processes. If a probability sample (e.g., a simple random sample) was not used, the basis for any inferences beyond the individuals in the sample must be discussed. Sensitivity to model formulation can be investigated by estimating the distribution of exposure for plausible alternative models if sample survey data have been collected for the input variables of the alternative models.

4.2 INTERVAL ESTIMATES OF EXPOSURE PERCENTILES: INPUT VARIABLE DISTRIBUTIONS KNOWN

When the joint probability distribution for the exposure model input variables can be represented by a specific family of probability distributions, confidence interval estimates for percentiles of the exposure distribution can be computed from a joint confidence region for the parameters of the input variable distributions. In general, let the vector $\underline{\theta}$ denote the parameters of the joint probability distribution for the model input variables. Let Θ denote a $100(1-\alpha)$ percent confidence region for the model input parameter $\underline{\theta}$, i.e.,

$$\text{Prob} \{ \underline{\theta} \in \Theta \} = 1-\alpha.$$

A $100(1-\alpha)$ percent confidence interval estimate of the p -th percentile of the exposure distribution is then given by

$$\{ Y_p(\underline{\theta}) \mid \underline{\theta} \in \Theta \},$$

where $Y_p(\underline{\theta})$ denotes the p -th percentile of the exposure distribution induced by the personal exposure model when the values of the input parameters are fixed at the value $\underline{\theta}$. That is, the set of all p -th percentiles, Y_p , of the induced exposure distributions generated by the set of $\underline{\theta} \in \Theta$ forms a $100(1-\alpha)$ percent confidence interval estimate of the p -th percentile of the population exposure distribution. Hence, the interval from the smallest to the largest

of these induced p-th percentiles, i.e.,³

$$[\text{Inf} \{ Y_p(\theta) \mid \theta \in \Theta \}, \text{Sup} \{ Y_p(\theta) \mid \theta \in \Theta \}],$$

constitutes an interval estimate of the p-th percentile of the exposure distribution having a confidence coefficient of at least $100(1-\alpha)$ percent. In practice, it is not feasible to derive the exposure distributions corresponding to all $\theta \in \Theta$. However, when the personal exposure model is mathematically tractable (e.g., continuous and monotonically increasing), it may be possible to deduce those values of $\theta \in \Theta$ that will generate the smallest and largest induced p-th percentiles. In this case, only these two estimates of the p-th percentile need to be computed in order to generate the confidence interval estimate.

Two advantages of this procedure with respect to that presented in Section 4.1 are that smaller sample sizes may be sufficient and that the confidence interval estimates should be somewhat narrower for a given sample size because additional information has been used (i.e., the assumed joint probability distribution of the input variables). A major disadvantage is that this procedure is more difficult to implement.

For example, suppose that an exposure model is based upon k independent input variables. When the input variables are statistically independent, the joint confidence level for the input variable distributions is the product of the individual input variable confidence levels. Thus, 95 percent confidence interval estimates for percentiles of the exposure distribution can be generated from confidence interval estimates of the parameters of the k independent variables, with the level of confidence being $(0.95)^{1/k}$ for the parameters of each input variable. If in addition each input variable distribution is indexed by ℓ parameters for which independent estimators exist, the joint $(0.95)^{1/k}$ confidence region for each input variable distribution is the cross product space defined by interval estimates for the ℓ individual parameters with a $(0.95)^{1/k\ell}$ confidence level for each interval.⁴ The resulting joint 95

³Technically, from the infimum (Inf) to the supremum (Sup) of the induced p-th percentiles corresponding to all values of θ in Θ .

⁴Confidence interval estimates of input variable parameters usually rely on a model, e.g., approximate Normal distribution for a sample mean. When $k\ell$ is large, this model may not be adequate for confidence interval estimates at the $(0.95)^{1/k\ell}$ confidence level.

percent confidence region for the k parameters of the input variable distributions generates a set of corresponding probability distributions for exposure via the model, based upon the k independent input variables. Given any particular set of input variable distributions, the corresponding distribution of exposures can be generated analytically or by using Monte Carlo simulation. For any given percentile of the exposure distribution, the 95 percent lower and upper confidence limits are the minimum and maximum, respectively, of the estimates of that percentile, based upon all exposure distributions generated by the joint 95 percent confidence region for the parameters of the input variable distributions (see, e.g., Section 4.2 of Appendix A).

These interval estimates for percentiles of the exposure distribution are a useful quantitative characterization of uncertainty for the exposure assessment. However, they are based upon many assumptions. Characterization of uncertainty for the exposure assessment is not complete without a thorough discussion of all assumptions. Justification for the model should be discussed as well as limitations of the data and design of the sample survey used to collect the data. If a probability sample was not used, any inferences beyond the sample subjects themselves must be justified. Any assumptions used in computing the confidence interval estimates, such as independence of model input variables or estimators of parameters of the input variable distributions, should be explicitly stated and justified. Sensitivity to model formulation can be investigated by estimating the distribution of exposure for plausible alternative models if sample survey data have been collected for the input variables of the alternative models. Using validated input variable distributions greatly reduces the uncertainty resulting from assuming particular input variable distributions. If exposure measurements are available, the personal exposure model should be validated also.

5. ASSESSMENTS BASED UPON EXPOSURE MEASUREMENTS

A major reduction in the uncertainty associated with an exposure assessment can be achieved by directly measuring the exposure for a sufficiently large representative sample of members of the affected population. This reduction in uncertainty is achieved by eliminating the use of a model to predict exposure. Moreover, a model can be validated and model parameters can be estimated if data are also collected for all potential model input variables.

The concept of directly measuring personal exposure to a substance is not new. For example, occupational exposures to radiation and carbon monoxide (CO) have been measured for years using personal exposure monitors (PEMs). However, PEMs which measure concentrations that are expected to be an order of magnitude lower than occupational concentrations have not previously been readily available. In the past decade, advances in technology have enabled manufacturers to miniaturize circuitry and components, thus making direct measurement of personal exposures much more feasible [24]. As a result, several personal exposure monitoring studies have been performed recently ([25], [26], [27]). These studies generally monitor an indicator of internal dose, such as continuously monitored concentrations in the contact zone, rather than internal dose per se.

Given a database of measured exposure values for individual population subjects, such as the hypothetical sample in Attachment A.2 of Appendix A, the population distribution of exposure may be estimated directly. In particular, the sample exposure distribution estimates the population distribution. If a probability sample design was used to select the sample members, a weighted estimate of the exposure distribution should be computed (see, e.g., sample survey inferences in [11], [21], [22], and [23]). Use of the sampling weights is especially important if high-risk segments of the population (e.g., individuals with potential occupational exposure, or individuals expected to be susceptible to health effects) have been oversampled. Typically, the sampling weights account for unequal probabilities of selection and weight adjustments to account for variations in response rate.

The measured exposure values may also be used to directly compute confidence interval estimates for percentiles of the exposure distribution (see, e.g., Section 10.18 of [22], or Section 12.9 of [23]). These confidence interval estimates for percentiles of the exposure distribution define an

approximate confidence region for the exposure distribution. When the exposure distribution is estimated from measured exposures for a probability sample of population members, confidence interval estimates for percentiles of the exposure distribution are the primary characterization of uncertainty for the exposure assessment. Design of the survey used to collect the data should also be discussed, as well as accuracy and precision of the measurement techniques. No mention of an exposure prediction model is necessary because the exposure levels were measured directly, rather than calculated from a model. If, however, the measured exposures are not based on a probability sample, it should be acknowledged that no strictly valid statistical inferences can be made beyond the units actually in the sample. If inference procedures are implemented for a nonprobability sample or if the sample design is ignored in the analysis, the assumptions upon which these inferences are based (e.g., treatment of the sample as if it was a simple random sample, or assumption of an underlying model) should be explicitly stated and justified [28].

When personal exposures are being monitored, the exposure distribution is often very skewed, with most of the sample members having exposure levels that are less than the detection limit of the monitoring equipment, especially when monitoring exposure to toxic substances ([27], [29]). In this situation, estimation of the distribution of exposure may require a very large sample size. Estimation of the proportion of population members with a detectable level in specified analysis subgroups and in the total population may then become an important analysis activity. Sometimes no sample members are found to have a detectable level. When no sample members have detectable levels, the power⁵ of the hypothesis test that the proportion of population members with a detectable level exceeds 1 percent, 2 percent, etc., may become the primary analysis [29]. It can then be inferred that the proportion of the population with a detectable level does not exceed some fixed percentage for which the power of the test is high. Qualitative characterization of uncertainty for the assessment would be basically unchanged, but the power of the hypothesis test for specified alternative hypotheses would be the primary quantitative characterization of uncertainty.

⁵The power of the test is the probability of correctly rejecting the null hypothesis that no population members have a detectable level when the alternative hypothesis (that the proportion of the population with a detectable level exceeds 1 percent, 2 percent, etc.) is true.

Whenever personal exposures are measured directly, quantification of measurement errors becomes an important aspect of the characterization of uncertainty. The terms "accuracy," "precision," and "bias" are used to describe specific aspects of measurement error ([5], [30], II.6-II.8 of [31]). The term "bias" refers to systematic errors, i.e., errors that have a nonzero expected value. Many chemical measurement techniques are known to produce biased measurements. For example, the amount of a contaminant in a specimen may be consistently greater than the measured amount. The term "precision" refers to the repeatability of a measurement. High precision means there is little variation between repeated measurements on the same object, i.e., the measurements exhibit a low coefficient of variation (ratio of the standard deviation of repeated measurements divided by the mean). Finally, accuracy refers to the cumulative effect of bias and precision. High accuracy occurs only when the bias term is small and the precision is high. Either high bias or low precision results in low accuracy.

When a survey collecting personal exposure data is properly designed, measurement error components can be quantified. Moreover, estimation procedures can be developed that compensate for the estimated measurement bias. A thorough quality assurance plan that enables estimation of measurement error components is presented in reference [32]. The design presented in reference [32] enables estimation of overall precision, bias, and accuracy of a parameter estimate via the relative standard deviation (RSD), relative bias (RB), and relative root mean square error (RMSE), respectively. These statistics are defined as follows:

1. The RSD is calculated as the ratio of standard error of a parameter estimate divided by the parameter estimate itself.
2. The RB is calculated as the ratio of the estimated absolute bias divided by the parameter estimate.
3. $RMSE = (RSD^2 + RB^2)^{\frac{1}{2}}$.

Another statistic that is sometimes used by the EPA as a measure of overall accuracy of a measurement procedure is the "total error" defined as the sum of the absolute bias plus twice the coefficient of variation. EPA guidelines specify that measurement accuracy is acceptable if the total error is less than 50 percent of the mean [33]. When the measurement error is unacceptable, improved measurement techniques are needed before any reliable analyses can be performed [33]. When the measurement error is acceptable, but

the measurement process has a nonzero bias, analysis techniques can compensate for the bias. Reference [32] discusses a survey design and estimation procedure that enables estimation of the mean exposure, corrected for the estimated mean bias. Reference [34] suggests that estimates of mean levels be multiplied by the relative bias to produce unbiased estimates of population means. Finally, reference [35] tabulates the true significance level of a two-sided hypothesis test as a function of the nominal significance level and bias for a survey design that does not allow estimation of the measurement bias. Uncertainty is then described qualitatively by saying that significant hypothesis test results would not be affected unless the bias was much greater than expected.

If previous research and/or theoretical derivations suggest that the exposure distribution should be a particular statistical probability distribution (e.g., Uniform, Normal, or Lognormal), the monitoring data may be used to estimate the parameters of this distribution. A measure of central tendency (e.g., mean, median, or mode) together with a measure of variability (e.g., standard deviation or a percentile) is usually sufficient to estimate such a distribution. When the statistical probability distribution has limited theoretical range, such as the Uniform distribution, the sample size and the smallest and/or largest data values are usually needed in order to estimate the extreme values of the distribution when they are not known a priori. Reference [18] provides formulas for estimation of the parameters of most commonly used probability distributions. When a theoretical probability distribution has been fit to the data, a statistical test of goodness of fit should be performed to determine whether or not the theoretical distribution provides a reasonably good fit to the monitoring data (see, e.g., Chapter 8 of [15], or Section 4.5 and Chapter 6 of [16]).

When total personal exposure via a particular route, e.g., inhalation, is monitored, the data provide a direct estimate of the distribution of total exposure for members of the sampled population. Since aggregation of estimated exposure distributions over sources and pathways is not required, the monitoring data provide a more defensible estimate of the distribution of exposure than the model-based estimates discussed in Chapters 2, 3, and 4. Models that link observed exposure to emission sources may be needed, however, in addition to monitoring data, in order to estimate the risk associated with regulatory options. If data were collected for potential model input variables as well as exposure (see, e.g., the hypothetical example in Attachment

A.2 of Appendix A), then the parameters of a model relating exposure to the potential input variables can be estimated using statistical regression techniques. Moreover, lack of fit can be tested for a hypothetical model, e.g., a model based upon the physical transport and fate process. Such models linking sources and exposures are very important for assessing the potential impact of regulatory actions. Using models validated in this way greatly reduces the quantitative uncertainty associated with exposure assessments. Unfortunately, the analytical measurement difficulties and costs associated with collection of directly measured exposure levels often severely limit or even prevent collection of such data.

6. AGGREGATING ACROSS SOURCES, PATHWAYS, AND/OR ROUTES OF EXPOSURE

6.1 COMBINING ESTIMATED EXPOSURE DISTRIBUTIONS

An exposure assessment based upon directly measured exposure levels generally estimates the distribution of total exposure via a particular route (inhalation, ingestion, or dermal) for the subjects of a subpopulation as discussed in Chapter 5. However, an exposure assessment based upon a model usually estimates the distribution of exposure for subjects in a subpopulation for a particular combination of source, pathway, and route. Integration of the assessment over the sources and pathways is then needed in order to estimate the total exposure distribution for an absorption route.

If the index s , where $s=1, 2, \dots, S$, indexes the unique combinations of sources and pathways for a particular route, r , and subpopulation, t , then the route r exposure for the i -th subject in the subpopulation can be expressed as

$$Y_{r,t}(i) = \sum_{s=1}^S Y_{r,s,t}(i). \quad (7)$$

The distribution of $Y_{r,t}(i)$ can be derived analytically or by Monte Carlo simulation from the distributions of $Y_{r,s,t}(i)$. In either case, an important consideration is whether or not the S source-pathway combinations are independent of each other. If not, the assessor should determine by sensitivity testing whether the correlated combinations must be simultaneously modeled so as to incorporate the correlation. For example, if two pathways are strongly correlated because they originate from the same source, the two pathways would need to be treated simultaneously in the Monte Carlo simulation. The source value for one pathway would then always be equal to the source value for the other pathway for each iteration.

The total route r exposure for the i -th subject in subpopulation t can be computed from Equation (7) by summation over the independent source-pathway combinations indexed by s . The distribution of route r exposure may then be estimated by summing Monte Carlo simulations of the S independent source-pathway exposure distributions. Thus, the Monte Carlo method could be implemented by simultaneously generating values for the independent source-pathway combinations and their sum. The result would be estimated exposure distributions for the independent source-pathway combinations and their sum, the total route r distribution for subpopulation t .

In some cases, the risk due to exposure to a substance may be related to total exposure over all routes (e.g., when there is no risk for exposures below some threshold level). Estimation of total exposure is also necessary to judge whether regulation of a particular route of exposure is an efficient approach to controlling total population exposure. The total exposure experienced by the i -th subject in subpopulation t is then given by

$$Y_t(i) = \sum_{r=1}^3 Y_{r,t}(i) = \sum_{r=1}^3 \sum_{s=1}^S Y_{r,s,t}(i), \quad (8)$$

where r indexes the three exposure routes (inhalation, ingestion, and dermal).

In order to estimate the distribution of the total exposure, $Y_t(i)$, it would be important to determine whether or not the exposures via the three absorption routes were independent random variables. If not, the correlated routes would have to be treated simultaneously in a Monte Carlo simulation, as discussed above for correlated source-pathway combinations. When the summation is over independent routes, the Monte Carlo simulation values for the independent routes may be added to produce a simulated value for the total exposure. The simulated Monte Carlo distribution of total exposure can then be generated by a large number of repetitions of this simulation.

When the subpopulation distribution of exposure (either total exposure or route r exposure) has been estimated as described above, uncertainty concerning the individual route and/or source-pathway exposure distributions should be characterized as described in Chapters 3-5. In addition, these methods should be applied, to the extent possible, to the aggregated subpopulation exposure distribution. For example, if personal exposure has been measured directly for various routes and/or source-pathways for a sample of members of a subpopulation, then the aggregate exposure has also been measured directly, and the methods described in Chapter 5 can be used to compute confidence interval estimates for percentiles of the aggregated exposure distribution.

6.2 LIMITING DISTRIBUTIONAL RESULTS

When the total exposure given by Equation (7) or (8) is the summation of several terms, central limit theory provides important distributional results. In either case, the total exposure is written as the sum of the positive random variables $Y_{r,s,t}(i)$. Hence, if a central limit type of result applies to the variables $Y_{r,s,t}$, then the limiting exposure distribution is the Normal distribution as the number of terms in the summation in Equation (7) or (8)

becomes large. Thus, when the total exposure is the sum over several sources, pathways, and/or routes, and no one term has a variance that is large compared to the variance of the total, then the form of the distribution for total exposure for subjects in subpopulation t will be approximately Normal (see, e.g., Section 14-1.1 of [19]).

The mean and variance for the total exposure, $Y_{r,t}$, can easily be estimated from the mean, variance, and covariance for individual source-pathway combinations as follows:

$$E[Y_{r,t}(i)] = \sum_{s=1}^S E[Y_{r,s,t}(i)], \text{ and} \quad (9)$$

$$\begin{aligned} \text{Var}[Y_{r,t}(i)] = & \sum_{s=1}^S \text{Var}[Y_{r,s,t}(i)] \\ & + 2 \sum_{s=1}^{S-1} \sum_{s'=s+1}^S \text{Cov}[Y_{r,s,t}(i), Y_{r,s',t}(i)], \end{aligned} \quad (10)$$

where E , Var , Cov represent the population mean, variance, and covariance, respectively. Of course, when the source-pathway combinations indexed by s are assumed to be independent, the second term in Equation (10) is zero. Substituting unbiased or consistent estimates directly into Equations (9) and (10) provides corresponding unbiased or consistent estimates of the mean and variance of the total exposure, $Y_{r,t}$, for route r for members of subpopulation t . These estimates may be used to fit the limiting Normal distribution directly. In fact, any measure of central tendency (e.g., mean, median, or mode), together with any measure of variability (e.g., standard deviation or a percentile), can be used to estimate the limiting Normal distribution for total exposure.

The limiting Normal distribution for total exposure is attained as the number of terms in the Summation in Equation (7) or (8) becomes infinitely large. In practice, the number of terms will always be finite. The primary importance of the limiting Normal distribution is that the distribution for total exposure will be approximately Normal when there are several terms in the summation, none of which dominates the variability of the total exposure. In this case the means, variances, and covariances of the terms in the summation are most important because they are sufficient to determine the limiting distribution. Other shape characteristics of the exposure distributions for the individual terms are then less important.

Although the limiting distribution could be estimated directly when the total exposure is the sum of several terms, research into the rate of convergence to the limiting distribution would be needed for proper interpretation. Empirical research would be needed to determine when the number of terms in the summation was sufficiently large to use the limiting Normal distribution. The results would likely depend upon the sources and pathways being summed and their relative importance.

If Equations (9) and (10) are used to estimate the limiting Normal distribution for total exposure, adequacy of the approximation of the true exposure distribution by the limiting distribution should be discussed. An empirical study on convergence to the limiting distribution might be needed. If total exposure was measured for a sample of population members, these data should be used to test the adequacy of the limiting distribution. Characterization of uncertainty with regard to the distribution of total exposure would also need to address uncertainty with regard to the estimated mean and variance given by Equations (9) and (10). If possible, confidence interval estimates of the mean and variance should be derived. Sensitivity of the predicted exposure distribution to plausible ranges of values for these parameters should also be investigated. In addition, limitations of the data used to estimate means and variances or to validate the distribution of total exposure should be discussed.

7. COMBINING OVER DISJOINT SUBPOPULATIONS

After the total subpopulation t distribution of exposure has been estimated for every subpopulation, the sizes of the subpopulations should be estimated. The number of subpopulation members experiencing a particular level of exposure can be estimated from an estimate of the size of the subpopulation and an estimate of the proportion of subpopulation members experiencing that exposure level. Moreover, the proportions of the total population members in the individual subpopulations can be used to combine the estimated distributions for the subpopulations to produce an estimate of the distribution of exposure for the total population.

Uncertainty concerning the sizes of the subpopulations should be addressed by discussing limitations of the data and methods used to estimate the subpopulation sizes. Confidence interval estimates of the subpopulation sizes should also be produced whenever possible.

If the total affected population of N subjects is partitioned into T disjoint subpopulations that each contain N_t subjects, the estimated cumulative probability distribution function for total route r exposure is given by

$$\begin{aligned} F_r(y) &= \text{Prob}(Y_r \leq y) \\ &= \sum_{t=1}^T (\hat{N}_t / \hat{N}) \text{Prob}(Y_{r,t} \leq y) \\ &= \sum_{t=1}^T \hat{P}_t \hat{F}_{r,t}(y), \end{aligned} \tag{11}$$

where $\hat{N}_t$ and $\hat{N}$ are the estimated sizes of the t -th subpopulation and the total population, respectively, $\hat{P}_t$ is the estimated proportion of the population in subpopulation t , and $\hat{F}_{r,t}$ is the estimated cumulative probability distribution function for route r exposure in subpopulation t . If estimates of the route r subpopulation distribution functions, $\hat{F}_{r,t}$, are replaced by estimates of the distribution functions, $\hat{F}_t$, for the total over all routes, Equation (11) produces the estimated distribution function for total exposure over all routes for the total population.

When the distribution of exposure has been aggregated across subpopulations as described above, uncertainty concerning the individual subpopulation exposure distributions is characterized as described in Chapters 3-6. In

addition, uncertainty concerning the total population exposure distribution (either for route r exposure or total exposure) can be addressed in terms of percentiles of this distribution. The p-th percentile of the route r exposure distribution is that exposure level, y_p , such that

$$p = \text{Prob} (Y_r \leq y_p) = F_r(y_p), \quad (12)$$

where $F_r(y_p)$ is the cumulative distribution function for the route r exposure evaluated at y_p . Likewise, given any fixed exposure level, y_c , an estimate of the percentile, $\hat{p}_c$, represented by y_c is

$$\hat{p}_c = \text{Prob} (Y_r \leq y_c) = \hat{F}_r(y_c), \quad (13)$$

i.e., Equation (11) evaluated at y_c . The variance of this estimator can be expressed as

$$\text{Var}(\hat{p}_c) = \sum_{t=1}^T \text{Var}[\hat{P}_t \hat{F}_{r,t}(y_c)], \quad (14)$$

since the T disjoint subpopulations are necessarily independent. Equation (4) in Section 3.2 can then be used to consistently estimate the variance of the estimated percentile $\hat{p}_c$. The estimated subpopulation proportion, $\hat{P}_t$, and the estimated route r subpopulation exposure distribution, $\hat{F}_{r,t}$, may often be statistically independent (e.g., when Census data are used to estimate $\hat{P}_t$). In this case, the first-order Taylor series approximation for the variance of $\hat{p}_c$ (see Equation (4)) is given by

$$\text{Var}(\hat{p}_c) = \sum_{t=1}^T [\hat{P}_t]^2 \text{Var}(\hat{p}_{c,t}) + [\hat{p}_{c,t}]^2 \text{Var}(\hat{P}_t), \quad (15)$$

where $\hat{p}_{c,t} = \hat{F}_{r,t}(y_c) = \text{Prob} (Y_{r,t} \leq y_c)$, i.e., the estimated proportion of subpopulation t members with route r exposure not exceeding y_c . Using an approximate Normal distribution for the estimated percentile, $\hat{p}_c$, an approximate 95 percent confidence interval estimate of the percentile estimated by $\hat{p}_c$ is

$$\hat{p}_c \pm 2\sqrt{\text{Var}(\hat{p}_c)}. \quad (16)$$

If values outside the interval [0,1] result from this procedure, the interval may be truncated, making the confidence level conservative.

The greatest potential difficulty in application of Equation (15) is evaluation of $\text{Var}(\hat{p}_{c,t})$. If the route r exposure was measured directly for a simple random sample of the n members of subpopulation t, then this variance

can be estimated as follows

$$\text{Var}(\hat{p}_{c,t}) = \frac{(\hat{p}_{c,t}) (1-\hat{p}_{c,t})}{n} . \quad (17)$$

In general, the variance expression given by Equation (17) would be multiplied by a design effect (or variance inflation factor) to reflect unequal sample selection probabilities and/or differential adjustments to compensate for survey nonresponse (see sample survey analysis in [11], [21], [22], [23]).

8. SUMMARY

A summary of the primary methods recommended for characterizing uncertainty for exposure assessments was presented in Table 1. Virtually all exposure assessments except those based upon measured exposure levels for a probability sample of population members rely upon a model to predict exposure. The model may be any mathematical function, simple or complex, that estimates the population distribution of exposure or an individual population member's exposure as a function of one or more input variables. Whenever a model that has not been validated is used as the basis for an exposure assessment, the uncertainty associated with the exposure assessment may be substantial. The primary characterization of uncertainty is at least partly qualitative in this case, i.e., it includes a description of the assumptions inherent in the model and their justification. Plausible alternative models should be discussed. Sensitivity of the exposure assessment to model formulation can be investigated by replicating the assessment for plausible alternative models.

When an exposure assessment is based upon directly measured exposure levels for a probability sample of population members, uncertainty can be greatly reduced and described quantitatively. In this case, the primary sources of uncertainty are measurement errors and sampling errors. A thorough quality assurance program should be designed into the study to ensure that the magnitude of measurement errors can be estimated. The effects of all sources of random error should be measured quantitatively by confidence interval estimates of parameters of interest, e.g., percentiles of the exposure distribution. Moreover, the effect of random errors can be reduced by taking a larger sample.

Whenever the latter is not feasible, it is sometimes possible to obtain at least some data for exposure and model input variables. These data should be used to assess goodness of fit of the model and/or presumed distributions of input variables. This substantially reduces the amount of quantitative uncertainty for estimation of the distribution of exposure and is strongly recommended. It is recognized, however, that it may not always be feasible to collect such data.

REFERENCES

- [1] Callahan, M. A., Johnson, R. H., McGinnity, J. L., McLaughlin, T. G., Severn, D. J., and Slimak, N. W. (February, 1983). Draft Handbook for Performing Exposure Assessments. Prepared by the U. S. Environmental Protection Agency, Office of Health and Environmental Assessment.
- [2] Executive Office of the President, Office of Science and Technology Policy (May, 1984 Draft). Chemical Carcinogens; Review of the Science and Its Applications. Federal Register, Vol. 49, No. 100, 21594-21657.
- [3] National Research Council, Committee on Prototype Explicit Analyses for Pesticides (1980). Regulating Pesticides. National Academy of Sciences, Washington, D.C.
- [4] Calabrese, Edward J. (1982). The Role of Exposure Data in Standard Setting. Toxic Substances Journal, Vol. 4, 12-22.
- [5] U. S. Environmental Protection Agency (1978). Source Assessment: Analysis of Uncertainty-Principles and Applications. EPA-600/2-78-004u, PB80-131485.
- [6] Cox, D. C., and Baybutt, P. (1981). Methods for Uncertainty Analysis: A Comparative Survey. Risk Analysis, Vol. 1, No. 4, 251-258.
- [7] Miller, C. (1978). Exposure Assessment Modeling: A State-of-the Art Review. EPA-600/3-78-065.
- [8] Drake, R. L. and Barrager, S. M. (1979). Mathematical Models for Atmospheric Pollutants. Electric Power Research Institute.
- [9] Moschandreas, D. J. and Stark, J. W. C. (1978). The GEOMET Indoor-Outdoor Air Pollution Model. EPA-600/7-78-106, PB-285-706.
- [10] EPA Exposure Assessment Group (May 1984). Stochastic Processes Applied to Risk Analysis of TCDD Contaminated Soil: A Case Study.
- [11] Kalton, G. (1983). Introduction to Survey Sampling. Sage Publications, Beverly Hills.
- [12] Haemisegger, E., Jones, A., Steigerwald, B., and Thompson, V. (September 1984 Draft). The Magnitude and Nature of the Air Toxics Problem in the United States. USEPA Office of Air and Radiation, Office of Policy, Planning and Evaluation.
- [13] Johnson, R. H., Jr., Bernhardt, D. E., Nelson, N. S., and Calley, H. W., Jr. (1973). Assessment of Potential Radiological Health Effects from Radon in Natural Gas. EPA-520/1-73-004.
- [14] U. S. Environmental Protection Agency (1983). Human Exposure to Atmospheric Concentrations of Selected Chemicals. Office of Air Quality Planning and Standards.

REFERENCES (cont.)

- [15] Daniel, Wayne W. (1978). Applied Nonparametric Statistics. Houghton Mifflin, Boston.
- [16] Conover, W. J. (1971). Practical Nonparametric Statistics. John Wiley and Sons, New York.
- [17] Linstone, Harold A., and Turoff, Murray, eds. (1975). The Delphi Method: Techniques and Applications. Addison-Wesley, Reading, Massachusetts.
- [18] Johnson, N. L., and Kotz, S. (1970). Distributions in Statistics: Continuous Univariate Distributions. Houghton Mifflin, Boston.
- [19] Burrill, Claude W. (1972). Measure, Integration, and Probability. McGraw-Hill, New York.
- [20] Woodruff, Ralph S. (1971). Simple Method for Approximating the Variance of a Complicated Estimate. Journal of the American Statistical Association, Vol. 66, 411-414.
- [21] Kendall, Maurice G., and Stuart, A. (1968). The Advanced Theory of Statistics, Volume 3. Hafner, New York.
- [22] Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). Sample Survey Methods and Theory, Volume 1. Wiley, New York.
- [23] Kish, L. (1965). Survey Sampling. Wiley, New York.
- [24] Wallace, L. A., and Ott, W. R. (1982). Personal Monitors: A State-of-the-Art Survey, Journal of the Air Pollution Control Association, Vol. 32, 602-610.
- [25] Spengler, J. D., and Soczek, M. L. (1984). Evidence for Improved Ambient Air Quality and the Need for Personal Exposure Research, Environmental Science and Technology, Vol. 18, No. 9, 2681-280A.
- [26] Hartwell, T. D., Clayton, C. A., Michie, R. M., Whitmore, R. W., Zelon, H. S., Jones, S. M., and Whitehurst, D. A. (1984). Study of Carbon Monoxide Exposure of Residents of Washington, DC, and Denver, Colorado. EPA-600/S4-84-031, PB84-183516.
- [27] Pellizzari, E. D., Hartwell, T. D., Sparacino, C. M., Sheldon, L. S., Whitmore, R. W., Leininger, C., and Zelon, H. S. (1984). Total Exposure Assessment Methodology (TEAM) Study: First Season-Northern New Jersey. RTI/2392/03-3S. Research Triangle Institute, Research Triangle Park, NC.
- [28] Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983). An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys. Journal of the American Statistical Association, Vol. 78, No. 384, 776-793.

REFERENCES (cont.)

- [29] Mason, R. E., McFadden, D. D., Iannacchione, V. G., and McGrath, D. S. (1981). Survey of DBCP Distribution in Ground Water Supplies and Surface Water Ponds. RTI/1864/05-08F, Research Triangle Institute, Research Triangle Park, NC.
- [30] Eisenhart, C. (1968). Expression of the Uncertainties of Final Results. Science, Vol. 160, 1201-1204.
- [31] Federer, W. T. (1973). Statistics and Society: Data Collection and Interpretation. Marcel Dekker, New York.
- [32] Lucas, R. M., Pierson, S. A., Myers, D. L., and Handy, R. W. (1981). National Human Adipose Tissue Survey Quality Assurance Project Plan. RTI/1864/21-11. Research Triangle Institute, Research Triangle Park, NC.
- [33] Pierson, S., and Lucas, R. (1983). Hispanic HANES Pilot Study: Measurement of Volatile and Semivolatile Organic Compounds in Blood and Urine Specimens. EPA 560/5-83-001.
- [34] Hosenfeld, J. M., Lucas, R. M., et al. (1983). Lead and Cadmium Levels in Blood of Baltimore, Maryland, City Public School Teachers: World Health Organization/United Nations Environmental Programs. Midwest Research Institute, Kansas City, Missouri.
- [35] Lucas, R. M., Iannacchione, V. G., and Melroy, D. K. (1982). Polychlorinated Biphenyls in Human Adipose Tissue and Mother's Milk. RTI/1864/50-02F. Research Triangle Institute, Research Triangle Park, NC.

APPENDIX A
A HYPOTHETICAL EXAMPLE

APPENDIX A: A HYPOTHETICAL EXAMPLE

A.1 INTRODUCTION

The purpose of this appendix is to provide a simple hypothetical example of a sequence of exposure assessments with a characterization of uncertainty for each. The sequence begins with an assessment based upon limited initial data and ends with an assessment based on sufficient monitoring data to validate and/or empirically estimate a personal exposure model. This example is intended to clarify concepts presented in the text of this report. It is recognized that many exposure assessments may not relate directly to this example. Nevertheless, it is expected that this example will provide a useful guide to characterization of uncertainty for exposure assessments and clarification of concepts discussed in this report.

The example is based upon a recent case study by the EPA Exposure Assessment Group.¹ The case study was concerned with an assessment of human exposure to TCDD (specifically 2,3,7,8-TCDD) resulting from TCDD contaminated soil located in an unsecure disposal site such as a sanitary landfill. It was assumed that the site was not located on residential or agricultural property and that access was sufficiently restricted so that significant human exposure occurred only offsite. Two routes of exposure, dermal and ingestion, were considered for exposure resulting from deposition of dust or eroded soil on residential areas. For the purposes of this illustration, only the soil ingestion route will be considered. The EPA case study focused on the population of individuals exposed to one part per billion of TCDD in the residential soil. A more refined assessment would have accounted for variability in the residential soil contamination levels.

A.2 ASSESSMENT BASED UPON LIMITED INITIAL DATA

The initial exposure assessment for a substance may be based upon limited data for directly measured personal exposures and/or variables that can be used to predict personal exposure. Assessments based upon limited data for directly measured exposure levels will be considered first. Assessments based upon limited data for the input variables of a personal exposure model will then be discussed in Section A.2.2.

¹EPA Exposure Assessment Group (May 1984). Stochastic Processes Applied to Risk Analysis of TCDD Contaminated Soil: A Case Study.

A.2.1 Limited Data for Directly Measured Exposures

An assessment may begin with directly measured exposure levels for a small sample of population members. The exposure data may be extant data, e.g., data from medical records or routine monitoring data. A sample survey might be designed to compile extant exposure data for sample subjects. Selection of the initial sample might be purposive (e.g., a sample of individuals with specific health effects and potential exposure to the substance under investigation). If a potential health problem was discovered, a more refined assessment based upon a probability sample of population members would almost always be needed to enable risk management decisions.

When an assessment is based upon a limited database of measured personal exposures, a primary aspect of the characterization of uncertainty is whether or not the data result from a probability sample. Characterization of uncertainty for an assessment based upon a probability sample of directly measured exposures is discussed in Section 5. However, statistics calculated from a small probability sample generally have relatively large standard errors, i.e., a relatively high level of uncertainty.

If the measured personal exposure levels for a small sample of population members were not so low that health risks were known to be negligible, a more refined exposure assessment might be pursued. Two ways of refining the exposure assessment would be as follows:

1. Measure exposure levels for a large enough sample to make inferences with a predetermined level of precision using the methods discussed in Section 5.
2. Develop an exposure scenario and associated exposure prediction model to be used as the basis for an assessment as discussed in Sections 2.2 through 4.2.

Since the exposure variable needed for predicting risk is often not amenable to direct observation, the latter model-based approach is often pursued first, if not exclusively. Hence, the model-based approach to refining the assessment is presented first.

A.2.2 Limited Data for Model Input Variables

If limited initial data on exposure indicate that a substance may present a significant health concern, the exposure assessment may be refined by developing a personal exposure model based upon an exposure scenario that considers all potential sources of the substance that could result in contact with population members. The model could be very simple, possibly involving only

multiplicative factors, or be very complex. In the present context, a personal exposure model is any equation that expresses a population member's exposure as a function of one or more input variables.

For example, suppose that the following model was adopted to predict the dose of TCDD absorbed in a lifetime by the i -th population member for the EPA case study cited above:²

$$Y(i) = [SC(i)] [ED(i)] [IR(i)] [AF(i)] [\varepsilon(i)], \quad (1)$$

where Y = Dose in units of nanograms of TCDD absorbed in a lifetime,

SC = TCDD concentration in ingested soil
= 1 part per billion (by definition in the EPA case study),

ED = Exposure duration in units of days of contaminated soil ingestion per lifetime,

IR = Soil ingestion rate for that portion of the lifetime during which soil is ingested in units of grams of soil ingested per day,

AF = Absorption fraction in units of grams of TCDD absorbed per gram of TCDD ingested, and

ε = A positive random error term with an expected value of one.³

For an initial exposure assessment, only limited data would typically be available concerning the population distribution of model input variables. For example, the data may support estimation of no more than approximate ranges of values for the input variables. For the model (1), suppose that the following estimates of the ranges of the input variables are based upon very limited data and/or subjective considerations:⁴

²This model is adapted from that used in the EPA case study. The case study model included a potency factor so that risk was estimated rather than exposure. The model also standardized the exposure by dividing by the exposed person's body weight and lifetime. The body weight and lifetime factors are not presently considered for the sake of simplicity. Also, lifetime, body weight, and exposure duration may well be correlated, especially for short lifetimes, making the Monte Carlo simulation much more complex.

³The random error term represents the cumulative multiplicative effect of all unobserved or latent variables not explicitly included in the model.

⁴In practice, it may be relatively easy to collect data with which to estimate the population distribution of some model input variables, e.g., SC and AF for model (1). The distribution of other input variables, e.g., ED and IR , or their product--the total grams of soil ingested per lifetime, may be much more difficult to estimate. One may not be able to improve upon expert subjective estimates of the ranges of such variables.

ED: 0 to 900 days of contaminated soil ingestion per lifetime.

IR: 0 to 11 grams of soil ingested per day of ingestion.

AF: 0.80 to 1.00 gram of TCDD absorbed per gram ingested.

Such data would typically serve as the basis for performing a sensitivity analysis for the model. The purpose of the sensitivity analysis would be to identify influential model input variables and develop bounds on the distribution of lifetime absorbed dose. The sensitivity analysis might begin by estimating the maximum possible dose. In the worst case (ED=900, IR=11, AF=1.0), the expected lifetime dose would be 9,900 nanograms of TCDD. If this dose was not considered to be a potentially significant health concern, the exposure assessment might be terminated. If one were not confident that the predicted dose truly represented the worst case, further investigation might be desirable.

If the worst case did present a significant health concern, the sensitivity analysis might be pursued further. For example, one might use the model to estimate the range of dose that results as each input variable varies over its range of possible values, with the other variables fixed at their midrange, i.e.,

ED: Dose varies from 0 to 4455 nanograms as ED varies from 0 to 900, with IR = 5.5 and AF = 0.90.

IR: Dose varies from 0 to 4455 nanograms as IR varies from 0 to 11, with ED = 450 and AF = 0.90.

AF: Dose varies from 1980 to 2475 nanograms as AF varies from 0.80 to 1.00, with ED = 450 and IR = 5.5.

Since the predicted dose is most sensitive to changes in exposure duration (ED) and soil ingestion rate (IR), collection of additional data for these variables, or their product--the lifetime soil ingestion--would be warranted. Since the plausible variation in the absorption fraction has less effect on the predicted dose, less effort might be directed toward collecting data on this factor; it might even be reasonable to treat this factor as a known constant for subsequent modeling.

The sensitivity analysis can be enhanced by computing the predicted doses that result from all possible input variable combinations. If each input variable had only a finite set of possible values, the set of all possible combinations of the input variables could be formed, and the predicted dose could be computed for each combination. These predicted doses could be used

to form a probability distribution of dose by computing the relative frequency of occurrence of each dose or interval of doses. The resulting dose sensitivity distribution is equivalent to the population distribution of dose only if all combinations of the input variable values are equally likely to occur in the target population. The dose sensitivity distribution can also be derived for personal exposure models based upon continuous input variables with finite ranges by representing these variables by equally-spaced points. At the limit, as the equal spaces become infinitely small and the number of points becomes infinitely large, the procedure for computing the dose sensitivity distribution is equivalent to estimating the probability distribution of dose that results from statistically independent continuous input variables with Uniform distributions on the estimated ranges. For the present example, the dose sensitivity distribution was generated by the Monte Carlo method with 3,000 iterations using independent Uniform distributions for ED, IR, and AF on the ranges 0 to 900, 0 to 11, and 0.80 to 1.00, respectively.⁵ The resulting distribution of dose is shown in Figure A.1. Estimates of selected percentiles of this distribution are shown in Table A.1. Statistics based upon the dose sensitivity distribution are interpreted in terms of all possible input variable combinations, rather than the population distribution of dose. For example, the 95th percentile of the dose sensitivity distribution (approximately 6400 nanograms of TCDD) is an estimate of the lifetime dose exceeded by only 5 percent of the doses resulting from all combinations of input variable values.

The uncertainty for this initial exposure assessment would be characterized by describing the limitations of the data used to estimate plausible ranges of model input variables and by describing the basis for the exposure model. A discussion of the limitations of the model and justification for the model would be a major component of the characterization of uncertainty. Plausible alternative models should at least be discussed. Sensitivity to model formulation should be addressed by repeating the Monte Carlo simulation for plausible alternative models. The choice of a specific number iterations

⁵Monte Carlo simulations were first performed with 100, 1,000, 3,000, and 10,000 iterations. All simulations presented in this appendix are based on 3,000 iterations because there was little difference in estimates based on 3,000 and 10,000 iterations. The additional cost of 10,000 iterations did not appear to be justified.

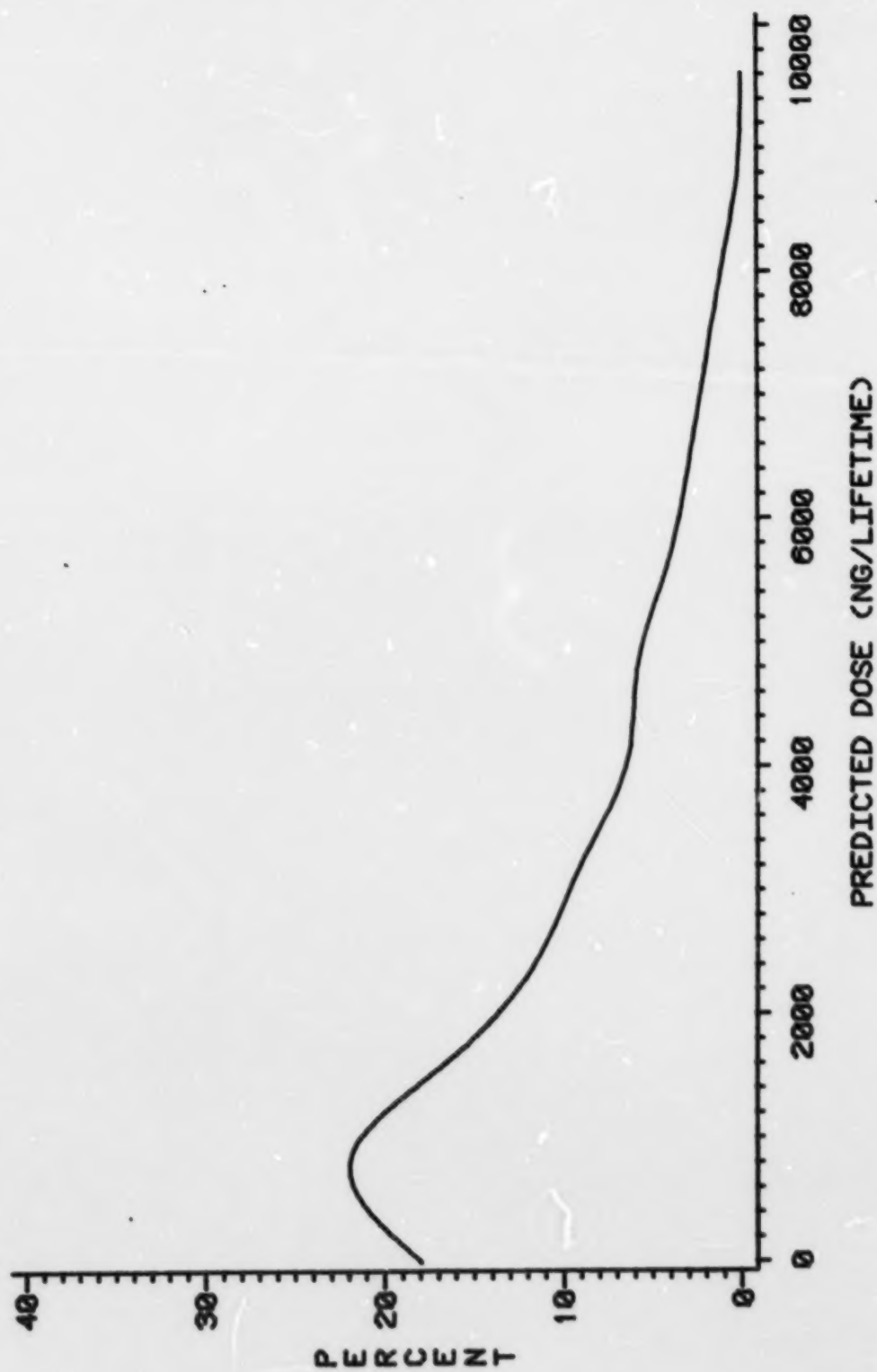


Figure A.1. Distribution of Estimated Doses Resulting from Equally Likely Input Variable Combinations

Table A.1. SELECTED PERCENTILES FOR ESTIMATED DOSE RESULTING FROM
EQUALLY LIKELY INPUT VARIABLE COMBINATIONS

Percentile	Point Estimate
0.01	11.57
0.02	26.15
0.03	39.64
0.04	59.55
0.05	78.44
0.06	103.72
0.07	128.05
0.08	151.81
0.09	172.36
0.10	194.40
0.15	314.61
0.20	467.12
0.25	616.11
0.30	790.89
0.35	995.34
0.40	1202.88
0.45	1408.98
0.50	1624.86
0.55	1909.83
0.60	2214.34
0.65	2571.43
0.70	2951.19
0.75	3355.48
0.80	3924.13
0.85	4552.30
0.90	5316.73
0.91	5473.16
0.92	5651.56
0.93	5865.06
0.94	6118.02
0.95	6379.43
0.96	6654.90
0.97	6939.50
0.98	7297.80
0.99	7828.36

Reproduced from
best available copy.

for the Monte Carlo simulation should be justified, noting that more iterations would improve the precision of estimates at somewhat greater expense.

A.3 ASSESSMENT BASED UPON ESTIMATION OF INPUT VARIABLE DISTRIBUTIONS

Discussions with subject-matter experts (e.g., medical doctors) can be used to formulate an estimate of the joint distribution of the model input variables. For example, suppose that discussions suggested that the input variables for model (1) were all statistically independent and that the following input variable distributions were supported:

ED: Discussions suggested that the distribution of exposure duration was a symmetric continuous distribution with the following properties:

1. Minimum duration = 0 days.
2. Maximum duration = 600 days.
3. Mean duration = 300 days.
4. 10 percent of exposure durations are 150 days or less.
5. 10 percent of exposure durations are 450 days or more.

This information suggests that the distribution of exposure duration is as shown in Figure A.2. This distribution can be parameterized by fitting a Beta probability distribution for ED/600.⁶ In particular, the Beta distribution with $\alpha = \beta = 3.1$ is a good approximation to the distribution of ED/600 shown in Figure A.2.

IR: Discussions suggested that the ingestion rate varied randomly from 0 to 5.0 grams per day over the population members, i.e., a Uniform (0, 5.0) probability distribution was supported for IR.

AF: Discussions suggested that the absorption fraction must be very close to one, if not identically one.

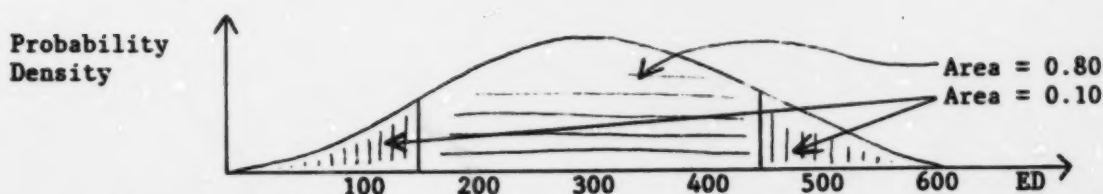


Figure A.2. Hypothetical Subjective Estimate of the Distribution of Exposure Duration (ED)

⁶See Johnson, N. L., and S. Kotz (1970). Distributions in Statistics: Continuous Univariate Distributions, Chapter 24 for the parameterization of the Beta distribution.

The estimated probability distributions for the model input variables were used to generate an estimated distribution of lifetime dose, assuming the above joint distribution for the input variables for model (1) using a Monte Carlo simulation based upon 3,000 iterations. The estimated distribution of dose is shown in Figure A.3. Selected percentiles from this distribution as shown in Table A.2.

The primary characterization of uncertainty for this exposure assessment would be a discussion of the justification for the presumed model and input variable distributions. Any available data would be used to validate the input variable distributions and predicted distribution of dose. Discussions with subject-matter experts are not sufficient to establish confidence interval estimates for percentiles of the estimated distribution of dose. However, differences in the expert opinions concerning the input variable distributions should be discussed. Moreover, sensitivity of the estimated distribution of dose to plausible alternative models should be discussed.

Comparison of Figure A.3 to Figure A.1 illustrates the effect of discussions with subject-matter experts for this hypothetical exposure assessment. Several changes occurred in the hypothetical example. The distribution of ED was changed from Uniform on the range (0,900) to a scaled Beta distribution on (0,600). The range of the Uniform IR distribution was modified from (0,11) to (0,5). Finally, the Uniform distribution of AF on the range (0.8., 1.0) was replaced with the fixed value unity. Such changes in the ranges of the input variables might occur if conservative estimates of input variable ranges were used for the initial exposure assessment. The net effect of all these changes on the worst case scenario is rather dramatic. The worst case scenario now predicts the maximum lifetime dose to be 3,000 nanograms of absorbed TCDD, instead of 9,900 nanograms.

When comparing Figures A.1 and A.3, another important change is that only Figure A.3 can be interpreted as an estimate of the population distribution of dose. It can be so interpreted because it is based upon estimates of the distributions of the model input variables. Figure A.1 was generated by treating all values within conservative estimates of the ranges of the model input variables as equally likely; it represents the dose sensitivity distribution, not an estimate of the population distribution of dose.

In order to isolate the effect of replacing the Uniform distribution for ED with a scaled Beta distribution, an additional simulation was performed

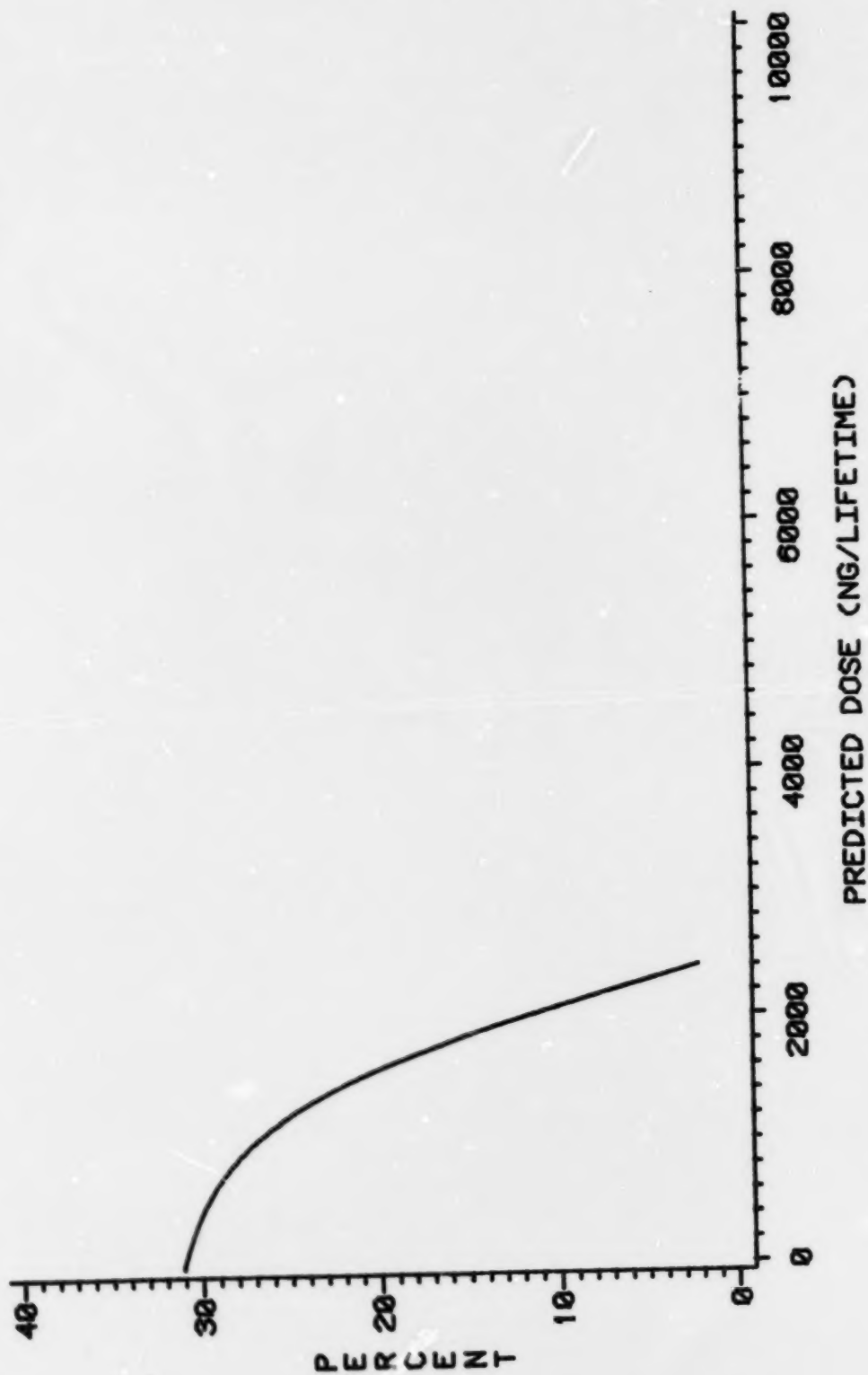


Figure A.3. Predicted Dose Distribution Resulting from the Presumed Model and Input Variable Distributions

Table A.2. SELECTED PERCENTILES FOR THE DOSE DISTRIBUTION RESULTING FROM
THE PRESUMED MODEL AND INPUT VARIABLES DISTRIBUTIONS

Percentile	Point Estimate
0.01	12.57
0.02	23.63
0.03	35.49
0.04	48.77
0.05	61.09
0.06	76.02
0.07	87.27
0.08	101.28
0.09	114.83
0.10	125.63
0.15	197.32
0.20	259.01
0.25	321.88
0.30	385.25
0.35	452.75
0.40	519.19
0.45	590.15
0.50	668.44
0.55	751.23
0.60	828.63
0.65	919.94
0.70	1019.62
0.75	1118.64
0.80	1244.86
0.85	1383.82
0.90	1547.47
0.91	1592.77
0.92	1626.87
0.93	1664.63
0.94	1721.11
0.95	1787.70
0.96	1842.86
0.97	1921.01
0.98	2034.71
0.99	2210.60

Reproduced from
best available copy.

using the same input distributions that produced Figure A.3, except that a Uniform (0,600) distribution was used for the ED distribution. Thus, Figure A.4 is based upon a Monte Carlo simulation with 3,000 iterations and independent input variables with the following distributions: (1) Uniform (0,600) for ED; (2) Uniform (0,5.0) for IR; and (3) $AF \equiv 1.00$. Selected percentiles from the resulting distribution of dose are presented in Table A.3 for comparison to Table A.2. Comparison of these tables reveals that the exposure distribution based upon a Uniform distribution for ED (Table A.3) is flatter and more disperse than that based upon the Beta distribution (Table A.2). This is a result of the fact that the Uniform distribution is itself flatter and more dispersed than the Beta distribution.

A.4 ASSESSMENT BASED UPON DATA FOR MODEL INPUT VARIABLES

The exposure assessment based upon an estimate of the joint probability distribution for model input variables can be refined by collecting sample survey data for all personal exposure model input variables for a probability sample of population members. The population distribution of exposure can then be estimated by computing the predicted exposure for each sample member based upon the model. These predicted exposures can be used to directly compute confidence interval estimates for percentiles of the exposure distribution. If the joint distribution of the model input variables is known, the sample survey data can be used to compute joint confidence interval estimates for the parameters of the input variable distributions, which can then be used to compute confidence interval estimates for percentiles of the exposure distribution. Using either method of computation, interval estimates for percentiles of the exposure distribution are a useful quantitative characterization of uncertainty for an exposure assessment.

Development of confidence interval estimates for percentiles of the exposure distribution would ordinarily require that sample survey data be collected for exposure and model input variables. That is, a probability sample of population members would be selected, and data would be collected for these individuals for the exposure and model input variables. For the example under consideration, the exposure of interest is the lifetime absorbed dose of TCDD via the ingestion route and the deposition pathway. The model input variables are the number of days of ingestion of contaminated soil in a lifetime, ED; the rate of soil ingestion, IR, in grams per day; and the absorption fraction, AF. Unfortunately, none of these variables are amenable to

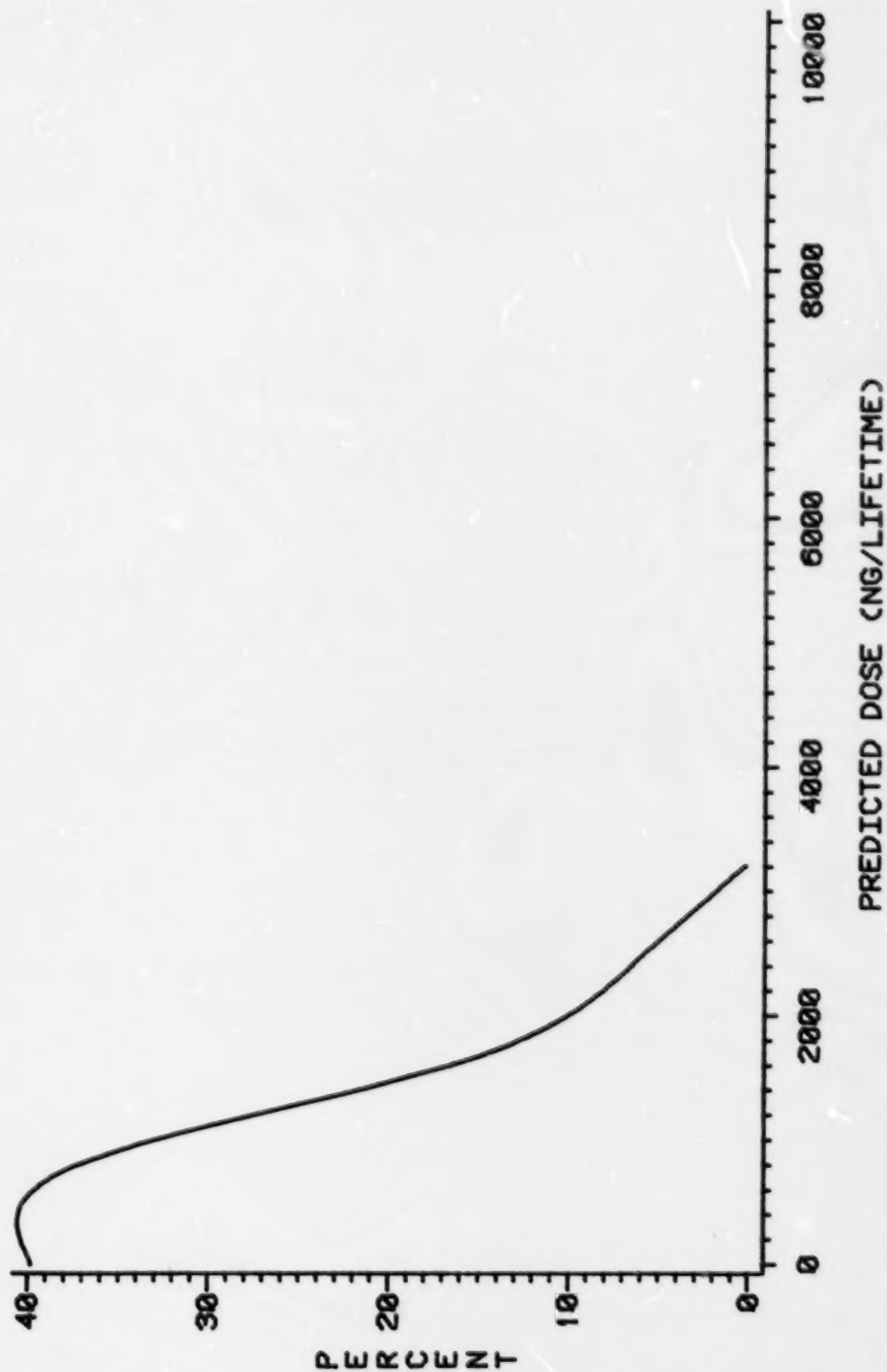


Figure A.4. Distribution of Estimated Dose Resulting from Alternative Equally Likely Input Variable Combinations

Table A.3. SELECTED PERCENTILES FOR ESTIMATED DOSE RESULTING FROM
ALTERNATIVE EQUALLY LIKELY INPUT VARIABLE COMBINATIONS

Percentile	Point Estimate
0.01	2.98
0.02	7.84
0.03	13.57
0.04	19.66
0.05	27.51
0.06	35.38
0.07	43.09
0.08	49.62
0.09	56.27
0.10	64.57
0.15	101.48
0.20	154.39
0.25	206.81
0.30	268.19
0.35	331.07
0.40	401.69
0.45	464.14
0.50	552.18
0.55	645.08
0.60	728.24
0.65	849.74
0.70	979.36
0.75	1134.78
0.80	1314.18
0.85	1503.40
0.90	1770.46
0.91	1842.22
0.92	1915.40
0.93	1980.76
0.94	2079.25
0.95	2158.24
0.96	2256.32
0.97	2328.70
0.98	2457.55
0.99	2628.13

Reproduced from
best available copy.



direct observation. For the sake of illustration only, this example will be continued as if the absorption fraction were known to be unity and the other variables could be measured directly for a sample of individuals from the target population.

A.4.1 Interval Estimates of Exposure Percentiles: Input Variable Distributions Not Known

For many personal exposure models, it may be much easier and less costly to collect data on the model input variables, rather than on the exposure itself. Hence, given model (1), suppose that data have been collected for the exposure duration, ED, and the ingestion rate, IR, for a probability sample of subjects from the target population. For illustration, hypothetical outcomes for a simple random sample of size 500 were generated using a symmetric Beta ($\alpha = \beta = 3.1$) distribution for ED/600 and an independent Uniform (0, 5.0) distribution for IR. The 500 pairs of values generated for ED and IR are listed in Attachment A.1. Histograms of the ED and IR distributions for this hypothetical sample are shown in Figures A.5 and A.6.

The predicted dose was computed for each of the 500 sample members as the product of the ED and IR values shown in Attachment A.1 in accordance with model (1), assuming that AF is unity. Treating these 500 predicted doses as if they were measured doses, equal tail area 95 percent confidence interval estimates were computed for percentiles of the distribution of dose.⁷ These confidence interval estimates are listed for selected percentiles in Table A.4 and presented graphically in Figure A.7.

Computation of the interval estimates for percentiles of the distribution of dose shown in Table A.4 and Figure A.7 is based upon: (1) the exposure model (1); (2) sample survey data on the input variables (Attachment A.1); and (3) the assumption that the absorption fraction (AF) is unity. The interval estimates of percentiles of the distribution are a useful quantitative characterization of uncertainty for the exposure assessment. However, characterization of uncertainty for the exposure assessment is not complete without a thorough discussion of all assumptions, especially justification of the model used to compute the predicted doses. The design of the sample survey used to collect the data should also be discussed. Moreover, sensitivity to the

⁷See Hansen, M. H., W. N. Hurwitz, and W. G. Madow (1953), Sample Survey Methods and Theory, Section 10.18; or Kish, Leslie (1965), Survey Sampling, Section 12.9.

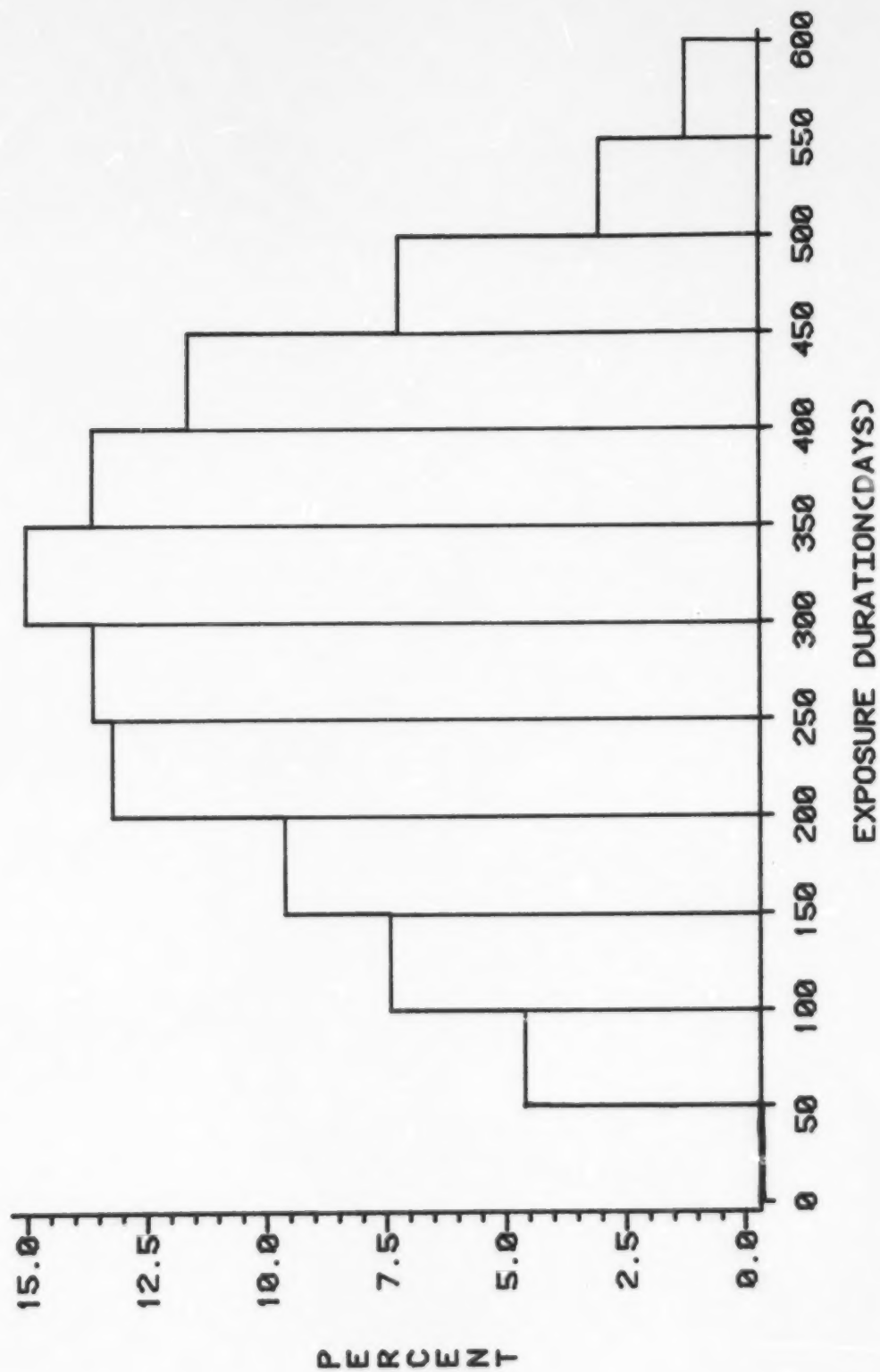


Figure A.5. Histogram of the Hypothetical Sample of Size 500 for Exposure Duration (ED)

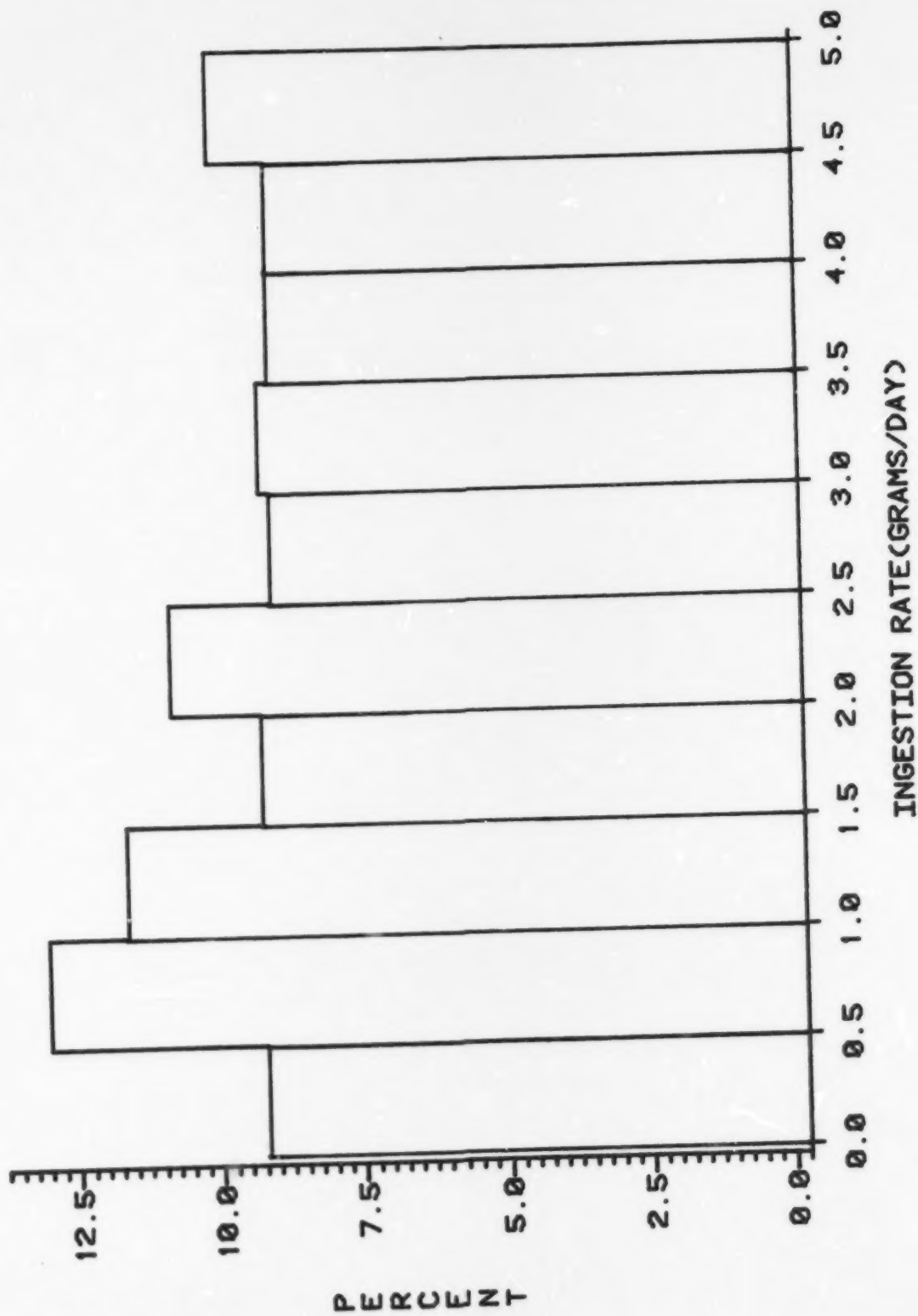


Figure A.6. Histogram of the Hypothetical Sample of Size 500 for Ingestion Rate (IR)

Table A.4. 95 PERCENT CONFIDENCE INTERVAL ESTIMATES FOR SELECTED PERCENTILES
OF THE DISTRIBUTION OF DOSE (Y) BASED ON MODEL-PREDICTED DOSES

Percentile	Lower Limit	Point Estimate	Upper Limit
0.01	0.30	7.88	10.25
0.02	4.80	10.69	25.28
0.03	9.63	23.12	46.49
0.04	13.26	35.66	57.01
0.05	23.79	49.53	72.68
0.06	35.36	61.23	82.15
0.07	48.86	74.69	124.46
0.08	56.38	81.32	141.70
0.09	67.01	114.38	149.59
0.10	78.97	137.07	162.53
0.15	156.50	186.11	212.77
0.20	196.79	235.89	266.90
0.25	257.81	276.87	311.68
0.30	285.10	332.24	373.27
0.35	339.18	391.30	445.83
0.40	400.67	452.37	534.44
0.45	462.55	548.58	604.82
0.50	559.40	608.62	678.24
0.55	624.04	682.09	748.56
0.60	694.09	751.23	861.61
0.65	778.75	873.85	949.18
0.70	885.71	955.25	1020.47
0.75	981.62	1055.00	1135.79
0.80	1088.97	1173.99	1291.32
0.85	1243.58	1331.26	1428.95
0.90	1416.62	1550.54	1725.20
0.91	1443.15	1630.07	1779.50
0.92	1529.68	1668.86	1866.79
0.93	1599.17	1747.71	1885.84
0.94	1668.66	1847.38	1944.57
0.95	1749.26	1880.79	2018.62
0.96	1859.73	1941.24	2127.76
0.97	1897.96	2027.12	2171.92
0.98	2007.93	2142.18	2295.75
0.99	2143.15	2280.78	2628.30

Reproduced from
best available copy.

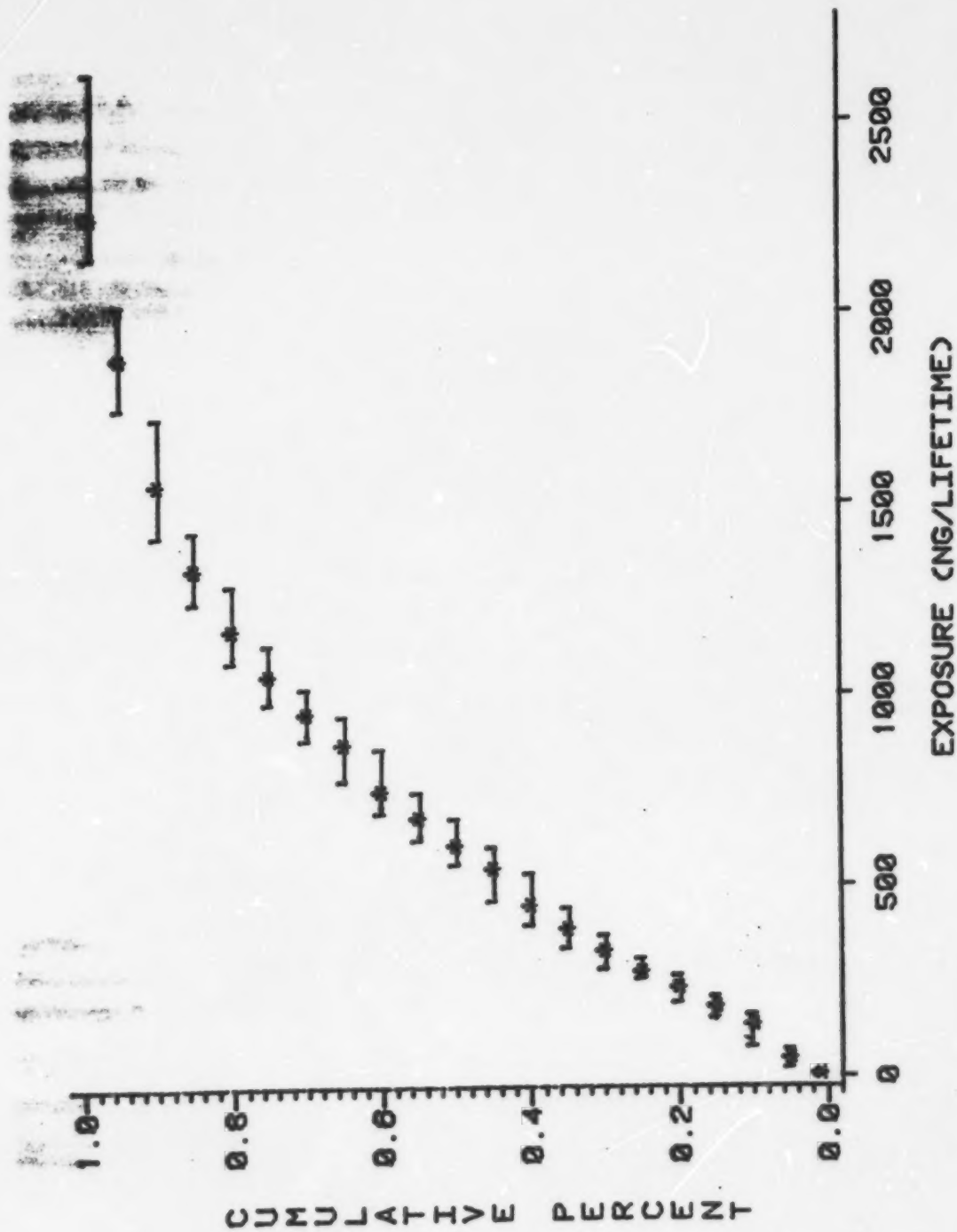


Figure A.7. 95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution Based on Model-Predicted Doses

assumed model should be investigated by computing the predicted doses and confidence interval estimates for plausible alternative models. If any directly measured exposure data are available (data on absorbed doses for the present hypothetical example), these data should be used to validate the predicted distribution of dose.

A.4.2 Interval Estimates of Exposure Percentiles: Input Variable Distributions Known

Confidence interval estimates were derived for percentiles of the exposure distribution in Section A.4.1 primarily by computing the doses predicted by model (1), given data for ED and IR. When the joint probability distribution for the personal exposure model input variables is presumed to be known, confidence interval estimates for percentiles of the exposure distribution can be generated from a joint confidence region for the parameters of the joint distribution of all model input variables. The level of confidence associated with interval estimates of exposure distribution percentiles is then the level of confidence associated with the joint confidence region for the parameters of the model input variables. When the input variables are statistically independent, the joint confidence level is the product of the confidence levels for the parameters of the individual input variables. For example, if the exposure model has only two input variables that are independent, then 95 percent confidence interval estimates of the exposure distribution percentiles are generated from the individual 97.5 percent ($\sqrt{0.95}$) confidence regions for the parameters of the two input variables.

Thus, for model (1), 95 percent confidence interval estimates for percentiles of the distribution of lifetime absorbed dose will be generated from 97.5 percent confidence regions for the parameters of the ED and IR distributions, assuming that the variables are statistically independent and that the absorption fraction, AF, is fixed at one. For the purpose of illustration, assume that expert opinion supports the Beta distribution as the true probability distribution for ED/600, and that the Uniform (0, θ) distribution is supported as the population distribution for IR. The data in Attachment A.1 will then be used to determine 97.5 percent confidence regions for the parameters of the ED and IR distributions in accordance with these assumptions. First, a 97.5 percent confidence interval estimate will be derived for the parameter θ of the IR distribution. Second, the data for ED will be used to

determine an approximate 97.5 percent joint confidence region for the parameters of the Beta distribution for ED/600. The 95 percent joint confidence region for the input variable parameters will then be used to generate 95 percent confidence interval estimates for percentiles of the exposure distribution.

Since the presumed probability distribution for IR is Uniform (0, θ), the Unique Minimum Variance Unbiased Estimator (UMVUE) for the parameter θ is⁸

$$\hat{\theta} = (n+1) IR_{\text{Max}} / n,$$

where IR_{Max} is the largest ingestion rate observed in a simple random sample of size n . Using this estimator, the UMVUE for the parameter θ based upon the data for IR in Attachment A.1 is 5.005. Moreover, the standard error (i.e., the estimated standard deviation) of this estimate can be shown to be

$$SE(\hat{\theta}) = IR_{\text{Max}} / \sqrt{n(n+2)} \doteq 0.010.$$

Since the maximum of a simple random sample of size 500 will have approximately a Normal probability distribution, an approximate equal tail area 97.5 percent ($\sqrt{0.95}$) confidence interval estimate of θ is from 4.982 to 5.027.

(Note that the true value θ , $\theta = 5.0$, which would not be known with real data, lies within this interval.)

The data for exposure duration (ED) in Attachment A.1 were used to compute maximum likelihood estimates of the parameters α and β for the Beta probability distribution for ED/600.⁹ The maximum likelihood estimates of these parameters were: $\hat{\alpha} = 2.97$ and $\hat{\beta} = 2.93$. A joint 97.5 percent ($\sqrt{0.95}$) confidence region for these parameters was then approximated by independent 98.7 percent ($\sqrt[4]{0.95}$) confidence interval estimates for $\hat{\alpha}$ and $\hat{\beta}$.¹⁰ Using

⁸See Patel, Jagdish K. (1973). Complete Sufficient Statistics and MVU Estimators. Communications in Statistics, Vol. 2, No. 4, pp. 327-336.

⁹See N. L. Johnson and S. Kotz (1970), Distributions in Statistics: Continuous Univariate Distributions, Section 24.4.

¹⁰The joint confidence level is 97.5 percent ($\sqrt{0.95}$) assuming independence, since the joint confidence level for independent estimates is the product of the individual confidence levels.

the limiting Normal distribution for these maximum likelihood estimates, the equal tail area 98.7 percent confidence interval estimates were computed to be from 2.54 to 3.40 for $\hat{\alpha}$ and 2.51 to 3.35 for $\hat{\beta}$. (Note that the true values, $\alpha = \beta = 3.10$, which would not be known with real data, lie within these intervals.) If the maximum likelihood estimators for $\hat{\alpha}$ and $\hat{\beta}$ were statistically independent, the cross product space defined by parameter values within the above ranges would be a joint 97.5 percent ($\sqrt{0.95}$) confidence region for these parameters. Instead, the estimators are positively correlated (correlation ~ 0.35), and the joint confidence region is shaped something like the ellipsoid shaded in Figure A.8. Since analytic specification of the 97.5 percent confidence region shown in this figure is very complex, the joint 97.5 percent confidence interval estimate for the Beta distribution parameters α and β was approximated by the cross product space defined by the interval estimates of the individual parameters, i.e., the rectangle in Figure A.8. This is a conservative estimate since the probability content of the rectangle is necessarily greater than that of the ellipsoid.

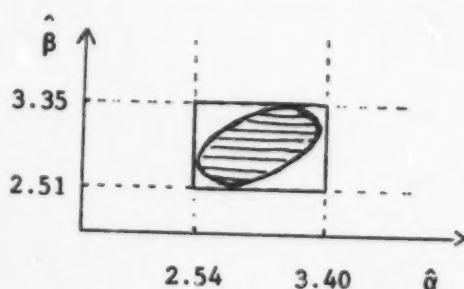


Figure A.8. Confidence region for the estimated Beta distribution parameters $\hat{\alpha}$ and $\hat{\beta}$

The cross product space defined by the 97.5 percent ($\sqrt{0.95}$) confidence interval estimate of the parameter θ of the IR distribution and the joint 97.5 percent confidence region for the parameters α and β of the ED distribution generates a corresponding set of dose distributions via model (1). Since the joint level of confidence associated with this cross product space is approximately 95 percent, an approximate lower (upper) 95 percent confidence limit for any percentile of the distribution of dose is given by the minimum (maximum) of that percentile for all dose distributions generated by input variable parameters within this cross product space.

Since the predicted dose is computed as the product of ED and IR for the model under consideration, the lower (upper) confidence limits for the dose distribution are generated by the ED and IR distributions that give the most probability to small (large) values of ED and IR. Hence, the lower limit of the parameter estimate for the IR distribution, $\theta = 4.982$, was used to generate lower confidence limits on the dose distribution, and the upper limit, $\theta = 5.027$, was used to generate upper confidence limits. Since the mean of the Beta distribution varies directly with α , the lower limit $\alpha = 2.54$, was used to generate the lower confidence limits for dose, and the upper limit, $\alpha = 3.40$, was used to generate upper limits. Since the parameter β affects mainly the variance of the Beta distribution for ED/600, it was necessary to consider both the upper and lower limits on β for generating each set of dose confidence limits, both lower and upper. Hence, the following procedure was used to compute the 95 percent confidence limits on the distribution of dose. To generate the lower confidence limits, two Monte Carlo simulations were performed as follows:

1. The first simulation consisted of 3,000 iterations computed using model (1) with the following inputs: AF = 1.0; IR distributed as a Uniform variate with $\theta = 4.982$; ED/600 distributed as a Beta variate with $\alpha = 2.54$ and $\beta = 2.51$.
2. The second simulation consisted of 3,000 iterations computed using model (1) with the following inputs: AF = 1.0; IR distributed as a Uniform variate with $\theta = 4.982$; ED/600 distributed as a Beta variate with $\alpha = 2.54$ and $\beta = 3.35$.

Percentiles were estimated for each of the dose distributions generated by the Monte Carlo method. The lower 95 percent confidence limit estimate was the smaller of the two estimates produced in this way for each percentile. The same procedure was used to generate the upper 95 percent confidence limits for dose, except that $\theta = 5.027$ and $\alpha = 3.40$ was used for the Monte Carlo simulations, and the larger of the two estimates was used for each percentile.

These 95 percent confidence interval estimates are listed for selected percentiles of the distribution of dose in Table A.5 and presented graphically in Figure A.9. The point estimates of percentiles shown in Table A.5 and Figure A.9 were computed directly from the 500 doses predicted by model (1) from the data in Attachment A.1. Thus, the point estimates in Table A.5 are identical to the point estimates shown in Table A.4. The confidence interval estimates are similar, but not identical, because they are computed using

Table A.5. 95 PERCENT CONFIDENCE INTERVAL ESTIMATES FOR SELECTED PERCENTILES OF THE DOSE DISTRIBUTION (Y) ASSUMING UNIFORM AND BETA DISTRIBUTIONS FOR IR AND ED/600, RESPECTIVELY

Percentile	Lower Limit	Point Estimate	Upper Limit
0.01	4.53	7.88	20.39
0.02	12.72	10.69	40.16
0.03	21.22	23.12	53.15
0.04	31.21	35.66	72.63
0.05	42.48	49.53	90.38
0.06	53.09	61.23	101.55
0.07	60.77	74.69	117.74
0.08	70.27	81.32	135.16
0.09	80.10	114.38	150.55
0.10	90.51	137.07	167.92
0.15	136.23	186.11	255.08
0.20	175.91	235.89	328.73
0.25	222.60	276.87	397.25
0.30	278.46	332.24	476.95
0.35	330.77	391.30	554.75
0.40	390.97	452.37	639.71
0.45	451.73	548.58	728.66
0.50	515.17	608.62	812.43
0.55	583.62	682.09	900.41
0.60	655.95	751.23	994.08
0.65	741.83	873.85	1098.52
0.70	822.19	955.25	1204.34
0.75	931.30	1055.00	1309.51
0.80	1045.89	1173.99	1445.75
0.85	1160.18	1331.26	1590.58
0.90	1329.38	1550.54	1782.87
0.91	1371.03	1630.07	1816.17
0.92	1399.65	1668.86	1863.04
0.93	1444.76	1747.71	1914.19
0.94	1513.55	1847.38	1964.64
0.95	1565.18	1880.79	2038.44
0.96	1639.32	1941.24	2122.65
0.97	1734.95	2027.12	2224.31
0.98	1824.22	2142.18	2348.82
0.99	1983.04	2250.78	2486.36

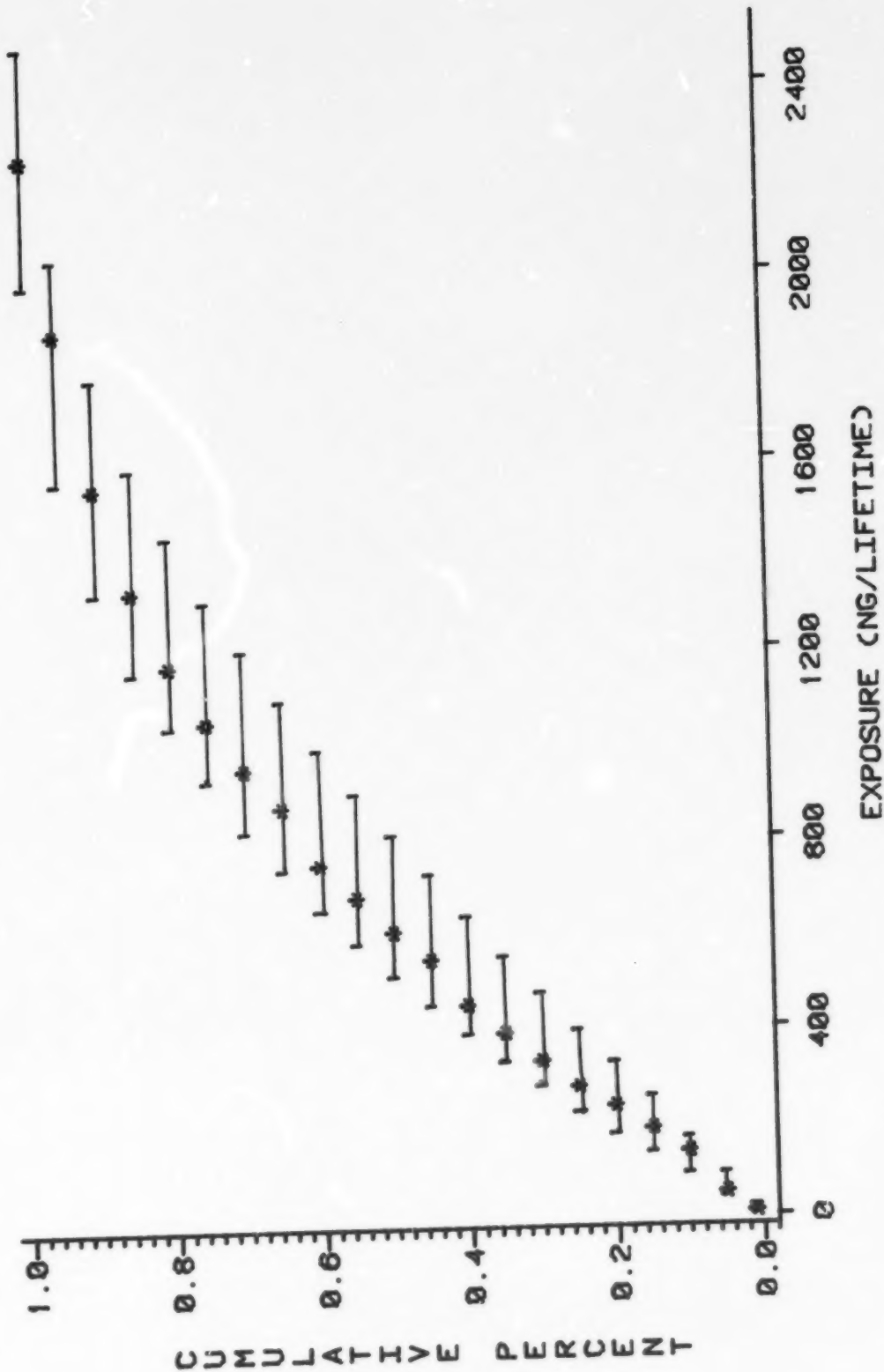


Figure A.9. 95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution Assuming Uniform and Beta Distributions for IR and ED/600, Respectively

different procedures, and the interval estimates shown in Table A.5 make use of the additional assumptions that a Uniform (0,0) distribution is accepted as the probability distribution for IR and an independent Beta distribution is accepted for ED/600. The interval estimates shown in Table A.5 should theoretically be narrower than those shown in Table A.4 due to the use of these additional assumptions. However, the opposite is true. This may be due to the conservative nature of the procedures used to estimate the joint confidence region for the parameters of the ED and IR input variable distributions. Moreover, the Normal distribution may not be a good approximation of the distributions of the parameter estimators ($\hat{\theta}$, $\hat{\alpha}$, and $\hat{\beta}$) when such high levels of confidence, 97.5 and 98.7, are necessary.

Computation of the interval estimate for percentiles of the distribution of dose shown in Table A.5 and Figure A.9 is based upon: (1) the exposure model (1); (2) sample survey data on the input variables (Attachment A.1); (3) the assumption that the absorption fraction (AF) is unity; (4) the assumption that the exposure duration (ED) and soil ingestion rate (IR) are independent random variables; (5) the assumption that the distribution of ingestion rate is Uniform (0,0); and (6) the assumption that the distribution of the scaled exposure duration variable ED/600 can be characterized by the Beta probability distribution. The resulting interval estimates for percentiles of the dose distribution are a useful quantitative characterization of uncertainty for the exposure assessment. However, characterization of uncertainty for the exposure assessment is not complete without a thorough discussion of the justification for all assumptions used in deriving these interval estimates. The assumption that a particular family of probability distributions can be used to represent a model input variable, e.g., the Beta family of distributions for ED, can be tested by standard goodness of fit tests.¹¹ Limitations of the data and design of the sample survey that produced the data should be discussed. Limitations of the procedures used to compute joint confidence interval estimates of the parameters of the input variable distributions should also be discussed. The number of iterations used in the Monte Carlo simulations should be justified, acknowledging that more iterations would produce more precise estimates at higher cost. Finally, sensitivity to the

¹¹See Daniel, Wayne W. (1978), Applied Nonparametric Statistics, Chapter 8; or Conover, W. J. (1971), Practical Nonparametric Statistics, Section 4.5 and Chapter 6.

assumed model should be investigated by computing the predicted doses and confidence interval estimates for plausible alternative models.

A.5. ASSESSMENT BASED UPON EXPOSURE MEASUREMENTS

The final stage of refinement for an exposure assessment is to collect sample survey data on exposure for a probability sample of population members. Such a database can greatly decrease the uncertainty associated with estimation of the population distribution of exposure if a sufficiently large sample is selected. This reduction in uncertainty is achieved mainly by eliminating the use of a model to predict exposure. Moreover, a model, which may be needed to relate exposures to sources, can be validated and its parameters can be estimated, if data are also collected for all model input variables.

Suppose that no alternatives to model (1) were under consideration, and data were collected for absorbed dose (Y), exposure duration (ED), and ingestion rate (IR), assuming that the absorption fraction (AF) was unity. For illustration, hypothetical outcomes for a simple random sample of size¹² 500 were generated using a symmetric Beta ($\alpha = \beta = 3.1$) distribution for ED/600 and an independent Uniform (0, 5.0) distribution for IR. Rather than assuming that model (1) predicts an individual's dose exactly, the dose, $Y(i)$, absorbed by the i -th population member was generated from the model

$$\ln [Y(i)] = 1.10 \ln [ED(i)] + 0.90 \ln [IR(i)] + \ln[\varepsilon(i)], \quad (2)$$

where $\ln$ denotes the natural logarithm, and the error term, $\ln(\varepsilon)$, follows a Normal probability distribution with a mean of zero and a standard deviation of one-third.¹³ Equivalently, the dose absorbed by the i -th population member was generated as

$$Y(i) = [ED(i)]^{1.10} [IR(i)]^{0.90} \varepsilon(i), \quad (3)$$

where ε is a random error term with a Lognormal distribution having a median of unity. The random error terms in models (2) and (3) can be thought of as the collective effect of minor explanatory variables not explicitly included

¹²In practice, the sample size needed to achieve acceptable precision within specified cost constraints must be derived. Sample sizes much larger than 500 may often be needed.

¹³It is recognized that data generated by this model are much simpler and more well-behaved than any real exposure data. Lacking real data with which to illustrate the proposed procedures, a simple hypothetical database has been generated.

in the model. The data generated for a hypothetical simple random sample of 500 members of the target population are presented in Attachment A.2.

Given a database of measured exposure values such as the hypothetical doses shown in Attachment A.2, the population distribution of exposure may be estimated directly. The sample exposure distribution is a direct estimate of the population distribution of exposure. The measured exposure values may also be used to directly compute confidence interval estimates for percentiles of the exposure distribution as shown in Table A.6 and Figure A.10.¹⁴ These confidence interval estimates for percentiles of the exposure distribution define an approximate confidence region for the exposure distribution. When the estimate of the distribution of exposure is based upon measured exposures for a probability sample of population members, confidence interval estimates for percentiles of the exposure distribution are the primary characterization of uncertainty for the exposure assessment. Limitations of the data and design of the survey used to collect the data should also be discussed. No mention of a model is necessary because the exposure levels were measured directly rather than calculated from a model.

If a database such as that shown in Attachment A.2 is available, the parameters of a model relating personal exposure to explanatory variables for which data were collected may be estimated using statistical regression techniques. Moreover, lack of fit can be tested for a hypothetical model based upon a postulated exposure scenario, such as model (1). Both of these techniques will be illustrated below.

Suppose that model (1), with an absorption fraction, AF, of unity, is expected to predict an individual's lifetime absorbed dose. This model can be expressed equivalently as

$$\ln Y(i) = \ln ED(i) + \ln IR(i) + \ln \epsilon(i), \quad (4)$$

where $Y(i)$ is the dose experienced by the i -th population member. It would then be reasonable to use the data to estimate the coefficients α_1 and α_2 of the regression equation

$$\ln Y(i) = \alpha_1 \ln ED(i) + \alpha_2 \ln IR(i) + \epsilon^*, \quad (5)$$

¹⁴See Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), Sample Survey Methods and Theory, Section 1.18; or Kish, L. (1965), Survey Sampling, Section 12.9.

Table A.6. 95 PERCENT CONFIDENCE INTERVAL ESTIMATES FOR SELECTED PERCENTILES OF THE DOSE DISTRIBUTION (Y) DIRECTLY FROM EXPOSURE DATA

Percentile	Lower Limit	Point Estimate	Upper Limit
0.01	8.46	20.30	31.70
0.02	19.74	34.18	78.86
0.03	23.96	78.18	97.05
0.04	58.31	90.82	128.52
0.05	78.39	105.47	137.30
0.06	90.06	129.99	151.25
0.07	100.43	137.32	191.73
0.08	127.78	149.44	206.15
0.09	134.52	181.19	228.47
0.10	143.55	204.65	258.13
0.15	239.87	319.20	371.08
0.20	354.23	394.30	480.99
0.25	414.71	520.98	598.39
0.30	534.03	622.84	678.32
0.35	636.54	703.66	804.16
0.40	720.13	830.11	912.92
0.45	848.09	932.71	1066.57
0.50	948.75	1099.23	1176.51
0.55	1101.77	1181.78	1353.50
0.60	1212.82	1363.13	1477.58
0.65	1377.45	1489.36	1633.91
0.70	1530.40	1690.40	1800.59
0.75	1716.53	1853.71	1982.21
0.80	1893.70	2075.07	2312.22
0.85	2202.19	2402.35	2667.51
0.90	2626.97	2966.36	3233.86
0.91	2727.32	3084.35	3305.52
0.92	2916.87	3169.80	3408.31
0.93	3081.00	3297.81	3592.00
0.94	3155.46	3337.69	3831.06
0.95	3298.06	3464.11	3889.16
0.96	3397.55	3809.79	4033.09
0.97	3672.42	3894.55	4253.92
0.98	3881.38	4092.06	4692.71
0.99	4106.65	4447.18	6466.43

Reproduced from
best available copy.



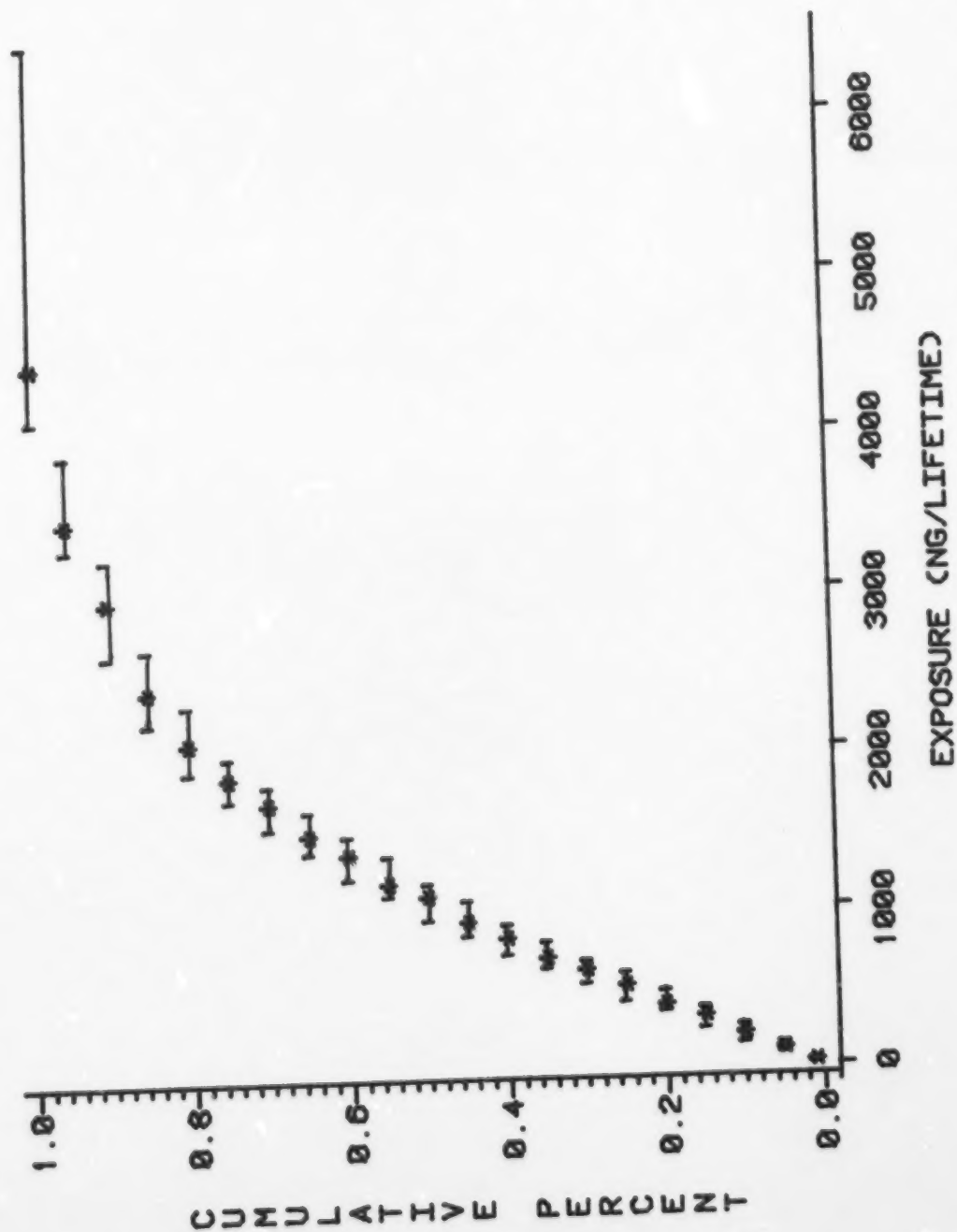


Figure A.10. 95 Percent Confidence Interval Estimates for Selected Percentiles of the Dose Distribution from Directly Measured Doses

where ε^* is a Normally distributed random error term with mean zero. Using the data in Attachment A.2, the estimated regression equation is

$$\ln Y(i) = 1.10 \ln ED(i) + 0.91 \ln IR(i). \quad (6)$$

The analysis of variance table for this regression equation is presented as Display A.1. The multiple coefficient of determination, R^2 , is a useful measure of goodness of fit of the model.¹⁵ The value of R^2 is very large for the present example ($R^2=0.998$) because the hypothetical data were generated from a model of the type being fitted. In practice, much smaller values of R^2 , e.g., $R^2 \geq 0.60$, may be regarded as indicating a reasonably good fit.

Having fitted the regression model (6), adequacy of the model (4) can be tested by testing the hypothesis that the regression coefficients, α_1 and α_2 , are both unity. In this case, the P-value of the hypothesis test is seen in Display A.1 to be essentially zero. That is, there is almost no chance that the estimated values of α_1 and α_2 can differ from unity by the observed amount or more when the true values of α_1 and α_2 are both unity. Thus, the hypothesis that model (1) generated the data can be rejected based upon this hypothetical sample of 500 measured exposures when the true state of nature is described by model (3). Of course, the true state of nature is never known for a real exposure assessment.

¹⁵The multiple coefficient of determination, R^2 , may be interpreted as the proportion of variability in the observed exposures explained by the model. It is the square of the multiple correlation coefficient, which measures the partial correlation between the dependent variable and the independent variables.

DISPLAY A.1: ANALYSIS OF VARIANCE TABLE FOR
REGRESSION EQUATION

DEP VARIABLE: LOG_DOSE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	2	23484.986	11742.493	112743.629	0.0001
ERROR	498	51.867778	0.104152		
U TOTAL	500	23536.854			
ROOT MSE		0.322726	R-SQUARE	0.9978	
DEP MEAN		6.774368	ADJ R-SQ	0.9978	
C.V.		4.76393			

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

VARIABLE	DF	PARAMETER	STANDARD	T FOR H0:		
		ESTIMATE	ERROR	PARAMETER=0	PROB > T	TYPE II SS
LOG_IR	1	0.911218	0.014461	63.010	0.0001	413.514
LOG_ED	1	1.102879	0.003020861	365.088	0.0001	13882.336

TEST: NUMERATOR: 70.8038 DF: 2 F VALUE: 679.8110
 DENOMINATOR: 0.104152 DF: 498 PROB >F : 0.0001

ATTACHMENT A.1: EXPOSURE DURATION (ED) AND INGESTION RATE (IR) FOR
A HYPOTHETICAL SIMPLE RANDOM SAMPLE OF 500 PEOPLE
FROM THE TARGET POPULATION

ID	ED	IR	ID	ED	IR	ID	ED	IR
1	290.65	3.92	51	375.17	3.73	101	218.54	3.56
2	214.40	1.23	52	447.16	0.35	102	262.14	2.40
3	408.64	0.43	53	170.43	1.55	103	179.78	3.77
4	447.18	1.43	54	265.41	4.96	104	319.49	3.75
5	250.06	2.99	55	266.33	2.55	105	73.64	1.36
6	188.59	2.59	56	390.30	2.19	106	347.19	0.24
7	182.29	0.78	57	187.03	1.99	107	120.68	1.63
8	420.17	1.44	58	67.41	4.03	108	171.22	3.56
9	459.12	4.99	59	261.90	1.26	109	75.28	1.07
10	450.48	0.63	60	366.93	2.32	110	248.33	4.93
11	155.22	3.61	61	392.56	0.43	111	282.11	0.03
12	300.34	4.19	62	290.85	2.64	112	413.10	1.04
13	418.41	0.42	63	162.69	3.74	113	314.39	3.18
14	396.53	0.70	64	133.24	0.24	114	221.86	0.77
15	228.30	0.76	65	339.34	0.53	115	382.24	3.93
16	314.87	4.48	66	442.77	2.27	116	384.95	1.79
17	562.12	2.49	67	124.05	3.60	117	270.34	2.39
18	329.69	1.13	68	308.89	0.75	118	555.35	1.90
19	369.26	4.73	69	275.64	3.44	119	272.88	4.25
20	168.81	3.53	70	102.18	2.71	120	144.87	2.78
21	74.99	2.99	71	362.25	3.21	121	415.69	2.14
22	406.58	2.02	72	285.35	3.85	122	169.05	0.24
23	273.09	0.89	73	262.76	1.73	123	297.44	0.15
24	330.12	2.62	74	179.92	1.60	124	346.39	4.59
25	193.45	2.39	75	313.86	0.16	125	155.87	2.63
26	252.09	4.36	76	58.94	4.50	126	489.48	1.47
27	407.47	0.40	77	238.38	0.48	127	233.19	4.01
28	314.19	2.91	78	120.16	3.25	128	502.69	2.41
29	297.10	0.06	79	223.07	3.75	129	165.89	3.39
30	140.30	1.42	80	305.42	1.41	130	364.77	2.90
31	263.93	3.72	81	397.14	1.51	131	73.24	1.92
32	358.40	1.48	82	279.84	2.13	132	407.04	4.93
33	420.39	2.69	83	85.86	4.65	133	176.11	1.01
34	374.70	0.40	84	470.18	4.56	134	333.77	3.52
35	345.24	3.84	85	211.76	3.43	135	332.18	4.60
36	361.91	4.21	86	506.76	4.17	136	83.97	3.66
37	203.61	3.43	87	377.40	4.77	137	389.57	4.28
38	148.00	4.39	88	150.65	2.97	138	177.97	1.29
39	306.31	2.23	89	166.30	1.71	139	520.51	0.64
40	488.84	1.44	90	260.47	3.37	140	203.79	0.17
41	480.23	0.01	91	407.41	4.98	141	279.42	3.59
42	446.08	4.27	92	427.99	0.84	142	211.96	3.39
43	300.99	2.90	93	367.61	2.89	143	325.04	0.01
44	94.15	0.37	94	371.64	3.44	144	350.04	4.15
45	266.08	0.03	95	244.73	1.93	145	398.25	0.67
46	351.15	0.81	96	299.93	2.44	146	142.18	0.33
47	398.21	3.38	97	260.08	0.72	147	172.45	3.89
48	437.03	2.90	98	124.43	1.03	148	223.76	4.79
49	52.99	4.86	99	452.48	2.04	149	147.22	3.89
50	115.67	2.69	100	255.49	3.75	150	363.94	3.44

**ATTACHMENT A.1: EXPOSURE DURATION (ED) AND INGESTION RATE (IR) FOR
A HYPOTHETICAL SIMPLE RANDOM SAMPLE OF 500 PEOPLE
FROM THE TARGET POPULATION**

ID	ED	IR	ID	ED	IR	ID	ED	IR
151	290.71	0.45	201	302.60	3.32	251	201.06	0.11
152	289.82	1.82	202	300.33	3.33	252	305.82	2.09
153	146.19	2.27	203	265.97	4.20	253	178.10	2.49
154	340.80	1.03	204	406.74	2.32	254	456.31	4.61
155	386.24	2.54	205	417.91	2.35	255	99.38	0.17
156	93.49	0.75	206	462.65	2.80	256	165.57	2.66
157	174.07	4.23	207	363.25	2.76	257	147.10	4.41
158	339.01	2.17	208	164.84	4.03	258	457.74	4.92
159	363.22	1.48	209	77.55	1.92	259	281.61	3.75
160	129.18	3.02	210	225.60	0.63	260	145.17	1.57
161	235.89	4.67	211	383.14	4.82	261	256.34	2.80
162	225.87	0.22	212	272.14	4.68	262	414.39	3.93
163	454.11	2.85	213	178.15	4.83	263	81.05	0.02
164	350.00	4.86	214	261.56	4.57	264	136.11	1.37
165	215.90	4.44	215	402.47	3.89	265	321.42	2.23
166	419.97	1.66	216	220.72	1.39	266	190.25	3.42
167	212.31	3.78	217	218.50	3.18	267	457.18	0.81
168	487.82	1.65	218	404.38	0.79	268	336.72	3.13
169	453.14	0.68	219	220.06	3.84	269	200.71	1.95
170	418.65	1.79	220	173.30	2.73	270	233.29	0.53
171	70.77	0.42	221	246.62	1.51	271	248.02	1.07
172	315.04	1.74	222	426.92	3.16	272	310.59	0.64
173	434.19	2.84	223	310.91	3.04	273	450.13	0.82
174	272.28	0.20	224	357.61	0.56	274	231.73	4.73
175	246.69	1.10	225	120.56	4.83	275	422.53	3.91
176	429.04	4.38	226	294.92	1.21	276	336.88	1.18
177	240.99	1.52	227	253.80	0.58	277	252.25	2.98
178	177.64	1.95	228	445.66	2.78	278	217.90	2.55
179	460.08	3.80	229	235.69	0.62	279	247.93	0.26
180	493.06	3.68	230	238.56	1.00	280	108.27	3.73
181	116.97	1.77	231	212.74	3.73	281	160.15	1.47
182	408.29	1.64	232	420.27	4.45	282	355.67	0.94
183	267.15	4.48	233	415.19	0.71	283	164.51	2.77
184	217.45	4.10	234	419.03	4.56	284	315.77	2.77
185	331.73	4.97	235	389.04	1.12	285	382.70	1.15
186	151.02	0.55	236	410.50	2.05	286	320.61	0.58
187	513.42	4.59	237	198.42	4.49	287	507.27	2.24
188	518.43	0.37	238	202.41	4.89	288	229.08	2.11
189	297.31	2.86	239	316.82	3.54	289	303.06	1.20
190	363.26	1.45	240	272.80	4.84	290	409.41	1.51
191	366.65	1.89	241	276.48	0.81	291	151.94	1.27
192	360.98	0.44	242	409.55	4.28	292	214.78	4.45
193	142.15	3.43	243	511.30	4.40	293	243.43	0.95
194	304.00	0.88	244	385.33	3.11	294	105.67	0.15
195	426.30	2.40	245	425.32	2.58	295	149.50	2.84
196	232.08	1.93	246	379.07	2.51	296	253.28	1.02
197	179.83	1.62	247	311.55	1.37	297	340.74	1.78
198	333.64	2.78	248	351.74	2.35	298	332.71	0.79
199	342.78	0.60	249	331.04	1.90	299	288.48	4.96
200	272.40	4.77	250	273.31	3.72	300	383.47	0.42

**ATTACHMENT A.1: EXPOSURE DURATION (ED) AND INGESTION RATE (IR) FOR
A HYPOTHETICAL SIMPLE RANDOM SAMPLE OF 500 PEOPLE
FROM THE TARGET POPULATION**

ID	ED	IR	ID	ED	IR	ID	ED	IR
301	393.45	2.19	351	221.42	0.89	401	187.26	3.05
302	344.82	1.39	352	335.00	3.84	402	300.64	0.95
303	212.93	4.13	353	370.84	1.63	403	472.08	4.90
304	389.78	4.11	354	393.83	1.24	404	300.31	0.64
305	500.62	2.92	355	100.20	0.69	405	563.68	4.63
306	294.89	0.89	356	208.90	2.60	406	349.32	1.06
307	134.22	1.99	357	129.49	1.33	407	397.86	0.03
308	262.37	0.04	358	99.63	0.66	408	237.77	0.99
309	127.55	0.19	359	397.81	2.26	409	311.65	4.39
310	171.75	0.05	360	297.83	4.95	410	266.95	3.46
311	457.60	4.38	361	253.05	4.38	411	331.72	3.27
312	246.83	3.57	362	423.20	0.18	412	359.59	0.61
313	90.56	4.83	363	75.64	3.59	413	233.74	3.20
314	380.47	3.32	364	349.01	4.86	414	301.77	3.31
315	340.34	2.64	365	275.75	3.16	415	344.15	1.76
316	194.93	2.12	366	325.21	3.00	416	488.20	1.49
317	366.48	1.42	367	206.72	4.89	417	299.64	0.60
318	286.90	1.16	368	238.86	1.83	418	184.04	1.41
319	244.63	1.40	369	402.65	4.64	419	266.18	3.03
320	187.49	1.91	370	405.64	2.51	420	356.35	2.39
321	443.75	4.81	371	325.77	3.49	421	292.60	2.13
322	561.98	0.87	372	264.66	3.58	422	193.74	1.55
323	363.83	0.21	373	433.21	1.19	423	212.38	2.73
324	267.38	1.44	374	399.53	0.81	424	232.28	1.34
325	259.42	2.54	375	136.78	4.98	425	186.02	2.11
326	492.69	3.15	376	381.40	0.70	426	184.09	0.00
327	264.89	1.86	377	126.25	0.64	427	349.22	2.73
328	218.92	0.61	378	326.25	2.00	428	212.82	1.00
329	80.76	1.00	379	413.84	0.72	429	541.22	2.33
330	231.79	3.05	380	385.01	2.43	430	405.26	3.54
331	441.88	3.21	381	454.95	4.13	431	110.92	0.55
332	140.11	0.98	382	370.76	1.76	432	332.53	0.90
333	120.99	2.16	383	256.94	2.11	433	273.23	2.15
334	356.12	3.99	384	515.14	0.31	434	453.08	4.85
335	314.07	3.03	385	438.94	3.07	435	440.38	1.34
336	224.59	1.41	386	427.27	1.39	436	178.93	3.36
337	477.29	4.25	387	245.59	0.78	437	276.17	2.43
338	293.35	0.57	388	542.53	1.09	438	424.58	0.35
339	117.50	4.52	389	536.84	3.99	439	307.01	1.01
340	470.54	0.72	390	488.31	1.63	440	319.32	1.46
341	175.42	3.27	391	354.61	0.70	441	509.45	0.89
342	421.65	4.15	392	384.50	1.68	442	468.82	0.43
343	388.37	2.58	393	396.62	4.89	443	73.63	4.06
344	395.61	1.81	394	454.49	4.08	444	266.12	4.47
345	458.83	2.44	395	308.33	4.50	445	94.49	0.59
346	266.76	2.20	396	462.45	0.56	446	302.46	2.38
347	377.46	4.99	397	273.44	3.56	447	299.70	4.75
348	232.55	1.12	398	228.67	4.46	448	94.19	1.94
349	335.38	4.06	399	327.20	1.13	449	303.70	1.66
350	463.88	2.42	400	214.14	1.99	450	361.50	2.98

ATTACHMENT A.1: EXPOSURE DURATION (ED) AND INGESTION RATE (IR) FOR
A HYPOTHETICAL SIMPLE RANDOM SAMPLE OF 500 PEOPLE
FROM THE TARGET POPULATION

ID	ED	IR
451	243.40	2.30
452	331.64	4.26
453	468.71	4.03
454	442.81	2.05
455	551.98	4.83
456	308.31	2.33
457	530.83	4.05
458	367.53	4.68
459	409.58	0.99
460	237.48	4.37
461	389.04	3.97
462	428.30	0.82
463	376.41	3.78
464	406.49	3.34
465	240.23	3.22
466	258.04	2.25
467	183.19	2.27
468	483.11	3.45
469	454.15	1.23
470	152.69	1.44
471	106.38	1.99
472	346.76	1.29
473	541.33	3.07
474	253.89	0.22
475	138.21	0.57
476	370.24	3.60
477	131.00	3.78
478	165.42	0.52
479	568.70	1.44
480	422.17	4.76
481	243.22	4.32
482	360.65	0.83
483	437.45	2.43
484	301.48	2.18
485	164.63	0.07
486	340.99	2.27
487	173.70	2.20
488	220.70	0.26
489	377.90	2.23
490	463.75	1.47
491	339.90	2.61
492	412.96	3.19
493	144.82	2.32
494	68.56	4.04
495	329.55	4.72
496	337.96	3.38
497	214.77	3.21
498	436.08	4.47
499	326.86	2.29
500	495.80	0.02

ATTACHMENT A.2: DOSE (Y), EXPOSURE DURATION (ED), AND INGESTION
RATE (IR) FOR A HYPOTHETICAL SIMPLE RANDOM SAMPLE
OF 500 PEOPLE FROM THE TARGET POPULATION

ID	ED	IR	Y	ID	ED	IR	Y
1	326.32	4.48	2610.66	51	314.01	2.52	1029.58
2	443.66	0.17	194.64	52	261.84	3.01	1159.51
3	545.18	4.47	3894.55	53	433.42	1.87	1230.45
4	411.15	4.79	2348.79	54	275.37	4.42	2509.21
5	219.71	4.93	2912.34	55	392.13	1.87	656.77
6	315.12	1.42	948.04	56	404.05	1.97	2628.06
7	369.06	0.43	533.20	57	524.02	3.05	1727.61
8	381.13	3.45	3792.96	58	328.67	3.11	2198.37
9	105.44	0.70	81.76	59	270.85	0.37	280.19
10	326.30	4.70	3736.52	60	271.41	3.74	1119.49
11	246.96	4.09	1676.04	61	228.39	0.37	96.13
12	434.19	0.04	34.18	62	357.70	1.85	1398.74
13	268.66	0.97	485.50	63	174.66	2.86	1123.23
14	389.97	4.65	2953.47	64	227.77	2.54	1382.99
15	397.93	0.65	459.32	65	422.48	2.55	1227.11
16	470.92	0.39	444.51	66	468.77	0.01	20.33
17	231.12	2.08	1631.28	67	333.58	1.02	877.52
18	434.74	2.16	1946.65	68	392.01	1.38	1099.60
19	370.77	1.08	620.83	69	168.53	2.70	748.63
20	126.59	4.50	861.69	70	165.56	0.23	97.63
21	237.30	4.62	1662.03	71	209.23	3.22	1146.43
22	418.19	4.40	3867.26	72	367.08	0.14	137.26
23	237.66	3.24	1163.74	73	209.99	1.42	481.58
24	198.72	2.85	653.99	74	523.47	4.81	3861.90
25	512.69	3.90	3388.27	75	88.54	0.79	88.96
26	304.95	3.97	1369.73	76	239.86	0.45	257.77
27	394.69	4.44	3874.05	77	292.97	3.78	2211.77
28	163.83	3.98	656.27	78	312.22	3.27	1889.22
29	241.84	4.11	1181.78	79	254.70	1.03	334.45
30	258.03	1.52	651.44	80	158.41	1.28	356.51
31	286.45	4.31	1123.49	81	426.11	1.86	932.71
32	241.74	0.06	24.91	82	315.50	4.20	1524.03
33	411.64	0.29	366.03	83	439.42	3.51	2293.45
34	24.22	0.24	8.46	84	378.92	1.27	702.49
35	241.80	3.32	1751.20	85	437.78	1.82	1281.00
36	324.04	0.67	544.04	86	119.15	1.60	241.08
37	170.87	4.18	1724.28	87	440.80	3.66	2019.67
38	370.58	3.10	3169.80	88	526.12	0.96	1241.00
39	210.09	4.62	1235.52	89	87.25	3.22	319.20
40	467.61	4.28	3124.42	90	368.36	0.18	164.36
41	389.61	3.31	2658.07	91	242.49	2.28	1444.50
42	334.57	3.81	1982.29	92	212.88	0.19	61.56
43	141.44	3.22	545.17	93	471.53	1.89	1730.98
44	316.61	0.31	215.78	94	180.26	0.35	129.15
45	346.67	4.56	3298.37	95	456.01	0.12	137.32
46	453.18	3.66	4011.03	96	474.83	0.06	76.91
47	306.73	2.31	1750.27	97	121.38	3.06	647.97
48	258.42	3.15	3101.61	98	406.53	0.51	660.62
49	325.82	2.81	1915.77	99	157.50	4.54	1470.68
50	487.37	3.84	2717.34	100	375.62	1.95	1179.37

ATTACHMENT A.2: DOSE (Y), EXPOSURE DURATION (ED), AND INGESTION
RATE (IR) FOR A HYPOTHETICAL SIMPLE RANDOM SAMPLE
OF 500 PEOPLE FROM THE TARGET POPULATION

ID	ED	IR	Y	ID	ED	IR	Y
101	293.50	3.29	1470.46	151	380.34	2.81	1985.66
102	266.37	3.10	873.52	152	251.88	4.38	1744.55
103	127.42	4.76	705.03	153	369.94	3.17	1808.72
104	317.70	3.25	1543.93	154	376.75	4.50	3589.79
105	288.04	3.05	2886.32	155	249.51	2.33	1123.10
106	414.07	1.56	913.83	156	424.62	0.93	1227.70
107	165.96	0.57	153.88	157	268.25	2.88	1604.09
108	105.28	0.88	127.61	158	305.95	2.89	4092.06
109	278.16	0.71	364.84	159	279.54	4.31	2135.47
110	344.82	2.63	3082.73	160	402.52	0.88	702.43
111	473.39	0.67	339.77	161	466.00	2.28	1979.11
112	192.95	0.03	16.48	162	77.26	2.82	405.48
113	169.41	1.80	329.64	163	230.33	3.16	1795.84
114	262.44	2.80	1381.76	164	528.49	2.18	1404.86
115	359.14	3.22	3247.36	165	422.08	0.82	971.49
116	246.12	0.21	141.49	166	178.33	1.36	219.50
117	293.88	2.36	1486.70	167	272.01	3.86	1180.76
118	92.97	4.70	639.93	168	564.25	3.02	3174.58
119	248.89	3.02	1140.40	169	366.62	1.57	931.79
120	462.41	2.52	2459.95	170	244.74	1.36	592.78
121	287.68	1.13	504.23	171	239.79	4.03	1504.52
122	358.09	2.11	1009.30	172	147.19	2.13	405.46
123	303.85	3.72	2167.90	173	294.51	3.51	1631.58
124	495.78	0.39	539.04	174	264.73	2.17	796.14
125	331.19	2.24	1691.81	175	204.42	3.21	1027.42
126	199.51	4.76	1489.36	176	536.87	4.57	4447.18
127	112.10	1.48	281.17	177	67.83	2.01	129.99
128	114.52	0.48	90.82	178	359.30	1.62	1163.06
129	192.99	2.55	860.89	179	523.41	2.33	2630.90
130	320.91	2.58	1252.56	180	353.22	4.09	2402.35
131	513.55	4.86	4853.12	181	292.50	1.01	480.32
132	258.19	4.57	2242.44	182	374.32	0.46	350.96
133	282.97	2.52	978.51	183	243.35	3.29	677.84
134	444.00	0.23	231.33	184	146.23	2.64	597.56
135	229.26	4.55	938.87	185	372.63	3.60	1944.95
136	297.87	0.14	106.08	186	364.34	4.10	2209.26
137	90.68	1.96	408.19	187	246.18	3.94	2223.72
138	401.26	1.10	592.93	188	289.84	0.95	378.65
139	384.72	3.27	1316.87	189	274.65	3.67	1139.50
140	342.52	4.74	3073.26	190	72.29	0.99	96.52
141	408.18	3.04	1433.05	191	359.75	4.30	1809.69
142	411.15	4.03	3297.81	192	306.52	0.84	192.47
143	205.20	2.45	1375.77	193	209.08	0.75	408.68
144	342.52	0.80	875.33	194	399.93	3.99	3300.37
145	143.56	4.55	1066.41	195	368.46	0.61	442.67
146	332.77	1.35	949.42	196	488.05	0.59	636.18
147	385.20	3.10	1856.78	197	324.87	2.69	1101.49
148	424.80	3.65	3179.39	198	256.81	4.95	1349.89
149	141.58	0.60	101.05	199	159.95	2.74	550.74
150	251.08	1.52	532.53	200	361.82	3.64	1787.22

ATTACHMENT A.2: DOSE (Y), EXPOSURE DURATION (ED), AND INGESTION
RATE (IR) FOR A HYPOTHETICAL SIMPLE RANDOM SAMPLE
OF 500 PEOPLE FROM THE TARGET POPULATION

ID	ED	IR	Y	ID	ED	IR	Y
201	362.44	3.83	3104.71	251	390.74	3.25	731.86
202	277.65	3.66	1604.86	252	435.27	1.05	625.68
203	91.10	2.50	414.59	253	394.80	0.42	358.61
204	462.02	2.47	1898.77	254	391.30	1.14	1054.86
205	330.27	4.30	1381.30	255	335.28	3.32	3601.92
206	354.09	2.36	645.84	256	494.75	1.39	1048.85
207	261.00	2.54	1627.68	257	258.81	3.59	1482.27
208	315.11	1.37	631.69	258	417.51	0.86	1702.38
209	485.20	0.38	305.79	259	442.75	4.68	3431.15
210	528.54	0.16	144.98	260	75.90	1.74	149.44
211	209.17	2.29	820.47	261	253.75	2.63	1224.78
212	146.49	3.32	725.38	262	449.49	4.20	4305.61
213	227.75	1.32	459.26	263	97.12	3.76	260.21
214	481.66	1.36	1351.95	264	268.94	4.78	1145.52
215	64.39	1.60	304.88	265	238.96	4.13	1386.73
216	397.65	3.50	1581.69	266	252.74	1.95	1377.04
217	199.70	2.99	859.10	267	247.35	1.05	534.11
218	408.30	3.78	2807.31	268	412.39	3.81	3969.85
219	491.48	0.41	380.24	269	238.33	0.93	668.53
220	266.18	4.95	1823.87	270	210.44	2.01	745.63
221	117.35	1.02	274.78	271	226.36	4.24	1464.93
222	277.33	2.42	847.11	272	261.31	0.17	22.92
223	487.47	0.48	850.87	273	345.39	1.97	1765.56
224	87.73	1.21	193.29	274	107.84	0.08	20.30
225	395.42	2.44	1882.56	275	381.64	0.57	393.06
226	477.10	2.26	1403.01	276	188.18	3.40	891.28
227	325.43	3.64	1596.33	277	394.49	1.52	2016.14
228	372.58	4.34	2612.63	278	292.27	2.94	988.57
229	522.04	0.66	553.37	279	368.73	3.96	2634.12
230	170.08	4.55	703.66	280	218.93	4.71	2601.51
231	371.03	3.34	1882.88	281	368.74	2.74	1759.68
232	170.02	0.06	30.30	282	115.92	2.73	674.39
233	288.47	1.66	959.89	283	230.48	3.81	598.40
234	531.10	4.23	5447.32	284	394.80	0.31	162.72
235	328.14	4.77	3410.80	285	231.38	0.60	181.19
236	402.86	0.49	396.64	286	340.71	4.53	1252.03
237	314.47	2.81	1023.19	287	344.18	1.07	1099.23
238	370.33	3.70	2169.70	288	367.29	4.75	1632.49
239	234.41	2.24	721.63	289	229.83	4.04	1712.29
240	268.85	0.79	496.02	290	230.20	0.67	317.87
241	413.74	3.88	2555.97	291	412.23	3.31	2415.30
242	524.16	1.64	1209.57	292	265.47	0.68	372.21
243	144.53	0.52	136.71	293	434.96	3.86	3809.79
244	140.67	1.51	394.30	294	141.84	3.61	951.92
245	248.26	0.95	560.04	295	329.09	2.87	2152.33
246	404.32	4.03	1989.41	296	248.40	4.37	1177.35
247	73.41	2.76	253.00	297	136.91	2.65	623.25
248	215.73	4.59	1699.06	298	232.17	2.44	706.76
249	160.98	3.07	805.07	299	294.16	3.94	1550.11
250	400.68	0.08	80.19	300	402.15	3.04	1690.40

ATTACHMENT A.2: DOSE (Y), EXPOSURE DURATION (ED), AND INGESTION RATE (IR) FOR A HYPOTHETICAL SIMPLE RANDOM SAMPLE OF 500 PEOPLE FROM THE TARGET POPULATION

ID	ED	IR	Y	ID	ED	IR	Y
301	272.22	4.43	1359.86	351	203.52	4.56	1353.88
302	246.97	0.53	209.72	352	399.24	4.53	4048.54
303	141.55	3.71	610.26	353	301.36	0.13	105.47
304	238.96	1.08	362.13	354	252.23	3.38	873.08
305	363.15	3.45	911.96	355	323.85	2.65	910.51
306	481.49	1.37	1333.77	356	383.55	4.08	3084.35
307	513.48	3.24	2144.92	357	279.64	2.20	1134.70
308	157.38	0.08	20.25	358	420.29	4.20	2851.95
309	436.84	3.93	3417.66	359	534.19	4.32	2327.10
310	368.75	4.53	3330.11	360	390.63	4.04	3032.80
311	51.17	0.53	56.32	361	224.80	4.49	1575.76
312	213.93	4.43	1083.00	362	313.91	3.29	2259.16
313	367.60	3.37	1634.06	363	356.01	2.92	1969.26
314	84.61	4.74	420.17	364	282.86	1.17	683.61
315	210.16	0.03	14.29	365	223.38	0.79	388.73
316	228.89	4.93	2305.25	366	356.07	2.23	1189.34
317	381.22	4.48	3134.67	367	475.03	2.94	2577.06
318	205.24	5.00	2978.33	368	194.84	3.84	1556.59
319	268.09	0.52	205.91	369	186.38	4.56	1423.10
320	252.40	1.17	555.72	370	371.97	3.18	1716.35
321	294.67	4.76	3337.69	371	236.69	2.28	813.47
322	445.65	4.58	6327.77	372	189.57	3.64	737.21
323	367.41	2.13	1125.09	373	347.26	2.04	1215.70
324	187.41	0.66	204.65	374	346.52	4.72	2825.99
325	135.45	1.30	249.11	375	102.28	1.36	191.56
326	287.63	0.40	339.51	376	237.86	1.64	763.04
327	169.57	0.89	520.98	377	464.65	2.96	2075.07
328	231.35	1.74	1061.80	378	229.34	2.17	621.29
329	361.41	0.94	502.58	379	336.67	3.79	1011.65
330	283.82	4.34	1799.85	380	419.55	1.71	1651.50
331	214.22	1.03	420.46	381	366.95	0.97	449.56
332	217.73	2.50	869.48	382	146.70	3.35	699.92
333	241.64	2.47	872.65	383	347.54	0.10	78.18
334	176.79	3.03	780.04	384	480.76	2.26	1697.68
335	361.28	4.27	2553.86	385	410.10	1.76	1968.30
336	370.36	4.19	2359.03	386	339.69	0.48	277.42
337	428.61	3.47	2436.67	387	255.41	1.97	948.69
338	495.79	1.43	1954.07	388	306.05	4.25	2033.70
339	412.18	4.86	2320.88	389	165.96	2.80	664.29
340	232.01	4.37	1473.41	390	212.11	1.18	322.80
341	415.01	4.66	4206.94	391	425.58	4.40	1945.46
342	181.00	3.98	1357.23	392	177.68	0.59	149.11
343	260.35	2.13	1275.75	393	157.73	0.42	131.60
344	462.32	3.43	1772.88	394	129.03	3.47	441.66
345	389.91	0.23	143.91	395	197.43	1.11	354.92
346	202.69	3.76	1101.57	396	458.45	3.25	2318.38
347	374.13	3.91	4667.53	397	391.57	4.21	2338.16
348	371.09	1.83	1122.30	398	242.79	2.18	1530.98
349	194.91	3.89	1441.90	399	290.08	3.33	1401.40
350	404.97	1.14	330.11	400	291.17	1.69	600.00

ATTACHMENT A.2: DOSE (Y), EXPOSURE DURATION (ED), AND INGESTION
RATE (IR) FOR A HYPOTHETICAL SIMPLE RANDOM SAMPLE
OF 500 PEOPLE FROM THE TARGET POPULATION

ID	ED	IR	Y	ID	ED	IR	Y
401	130.05	1.36	279.25	451	406.48	0.24	167.21
402	463.65	4.32	2672.62	452	222.52	1.52	622.84
403	363.04	1.85	664.13	453	185.98	3.61	958.33
404	162.80	3.79	1102.59	454	451.45	3.27	1494.14
405	231.61	2.22	740.42	455	498.03	4.00	3882.53
406	398.86	4.23	3464.11	456	318.32	1.12	451.19
407	350.84	4.86	6712.26	457	218.15	3.81	838.56
408	341.64	1.37	795.13	458	394.50	2.00	2116.80
409	310.44	1.88	1167.71	459	203.44	1.87	511.77
410	168.68	2.50	566.13	460	230.89	2.61	1202.64
411	143.38	0.33	78.65	461	341.12	4.78	1066.61
412	147.65	2.63	536.29	462	565.79	1.69	1363.13
413	510.92	0.93	522.30	463	450.22	1.12	883.91
414	463.93	0.61	1298.93	464	417.28	4.16	2739.13
415	196.28	3.92	1560.18	465	402.41	0.60	225.08
416	234.43	1.87	804.33	466	153.55	3.96	794.61
417	295.37	1.92	910.38	467	326.39	2.30	1551.73
418	134.98	1.61	206.18	468	407.69	0.17	132.67
419	359.39	0.19	120.72	469	233.02	2.47	1374.64
420	246.79	1.52	794.57	470	388.56	4.66	4114.89
421	433.13	4.83	3309.87	471	217.49	1.66	353.62
422	144.24	4.84	984.89	472	271.53	1.21	391.48
423	490.79	3.84	4328.40	473	277.40	2.74	1522.62
424	124.65	3.82	676.88	474	79.94	2.13	373.08
425	478.38	2.99	3231.51	475	149.72	4.07	924.12
426	229.50	1.47	598.11	476	334.82	1.25	1096.14
427	253.53	0.12	61.16	477	477.41	4.04	2359.30
428	207.96	1.13	310.05	478	357.03	1.85	1129.83
429	450.47	0.37	239.21	479	230.56	2.27	739.54
430	234.93	4.74	1853.71	480	416.37	4.30	1862.69
431	314.81	1.81	746.78	481	292.80	2.39	1250.12
432	482.26	3.10	1883.23	482	539.96	1.25	1162.79
433	421.57	2.44	1947.29	483	163.29	1.05	375.09
434	413.33	3.64	3913.18	484	225.76	3.42	1421.20
435	491.64	0.56	499.52	485	173.63	4.87	1278.63
436	340.84	3.90	2102.63	486	359.00	0.85	614.05
437	144.60	4.42	718.44	487	249.81	3.36	866.27
438	151.51	1.63	437.32	488	386.77	4.23	1731.97
439	115.21	4.79	847.43	489	387.05	2.94	802.62
440	338.97	3.05	1821.90	490	321.11	0.35	220.23
441	207.69	2.55	792.93	491	448.26	2.13	1797.12
442	94.91	4.73	532.17	492	298.84	3.23	1952.85
443	115.82	2.34	363.40	493	294.68	4.68	1486.85
444	491.54	3.25	2115.23	494	199.81	2.89	850.76
445	371.86	0.62	368.97	495	201.62	4.07	981.88
446	240.09	4.62	1568.00	496	262.76	0.60	261.11
447	354.88	3.89	2620.69	497	182.83	3.96	649.69
448	415.86	0.60	380.07	498	172.19	4.08	1866.66
449	184.03	4.40	912.11	499	231.20	4.90	2287.70
450	69.88	4.75	640.30	500	214.95	4.94	2966.36